Understanding the practical utility of using the analytic potential of patient data in Identifying High-cost patients

Kevin Li

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2018

Committee:

Adam Wilcox

Thomas Payne

Program Authorized to Offer Degree:

Biomedical Informatics and Medical Education

University of Washington

Abstract

Understanding the practical utility of using the analytic potential of patient data in Identifying the High-cost patients

Kevin Li

Chair of the Supervisory Committee:
Adam Wilcox
Biomedical Informatics and Medical Education

It is widely known that the minority of patients make up the majority of healthcare costs. Research being done aims at identifying these patients through predictive modeling. In the hopes that providing targeted resources to these patients can prevent inurnment of the high-cost. Lowing the bottom line to the hospital and helping the patient. Yet what degree of utility do these models provide? Most models are applied in a less than realistic setting or fail to state which predicted patients can even be impacted. In this study, I went through patient's clinical notes to better understand how practical such predictive models are. First, I sought after literature to better understand what variables most predictive models use as a base. I compare these to what was available in the patient's profile. Then revise what necessary for me to predict high cost given the patient's clinical notes. With access to UWMC/Harborview and NW Hospital databases, I went through clinical notes to evaluate each patient's possible

predictability. These determinations were later verified by a physician for accuracy. This was further reflected on Northwest(NW) Hospital data, which is a relatively smaller hospital with a focus on inpatient/outpatient patients. Each patient was categorized on the nature of their high expenditure. This work's importance is in how to consider predictive models moving forward. Assuming modeling will always have the solution to predict high-cost patients is misguided. Instead, understanding the underlying dynamic of the patient's cause is a better target. The conclusions made in this study can help better guide models to be more cognizant in how they approach predicting high-cost patients.

Executive Summary:

A glaring health issue is that minority of patients in healthcare cover the majority of healthcare expenditures. In response to this widely held observation, much research has been done on models to identify these patients to some degree before they incur the cost for possible prevention measures. These models while having predictive accuracy don't perform well outside of the facilities where it was generated from. There hasn't been any model that has been used widespread as a result. This study aims to better understand the dynamics of why patients become high cost to help frame prediction models.

This was down through the analysis of 200 patients from both UWMC/Harborview and Northwest Hospital respectfully. These patients were taken from the top 10% of all patient expenditures from their respective hospitals. Data collection was done manually through patient clinical notes. Qualitative analysis was done to sort out patient's predictability for their high cost and category that their high cost could fit into. Results were verified by an external physician for accuracy.

The findings from this cohort showed that majority of patients are unpredictable. The patients improbability for prediction stems from the lack of previous patient data and the nature of the diagnosis. This prediction is further mirrored regardless of

hospital type at NW Hospital showing similar degrees of unpredictability in the majority of patients. Yet upon further inspection the circumstances of the unpredictability appear to be unique to each hospital.

Based on these findings, it is clearly apparent that predictiveness is throttled by the inherent qualities of the diagnoses. Models bypass this when they are able to collect data from networks with multiple hospitals, but this isn't realistic. Trying to achieve predictability is very hard and only a small fraction is actually predictable. Of which, the utility of some of the predictable patients is diminished.

Models are improving at an alarming rate, but for models to be feasible in more realistic conditions an approach to better understand the why patients are high cost than how they are predictable must be the focus.

Background/Significance:

How are U.S Health care expenses distributed:

It is reported by literature from many various sources that a relatively small proportion of patients consume the majority of healthcare resources. In 2014, the Agency for Healthcare Research and Quality (AHRQ) used data from the US Medical Expenditure Panel Survey(MEPS) to help illustrate and estimate the concentration of health care costs across the U.S civilian noninstitutionalized population(Cohen, 2012).

AHRQ found that the top 5% of the population accounted for around 50% of total expenditures and the top 10% covering almost two thirds of total expenditures. (as shown in Figure 1).

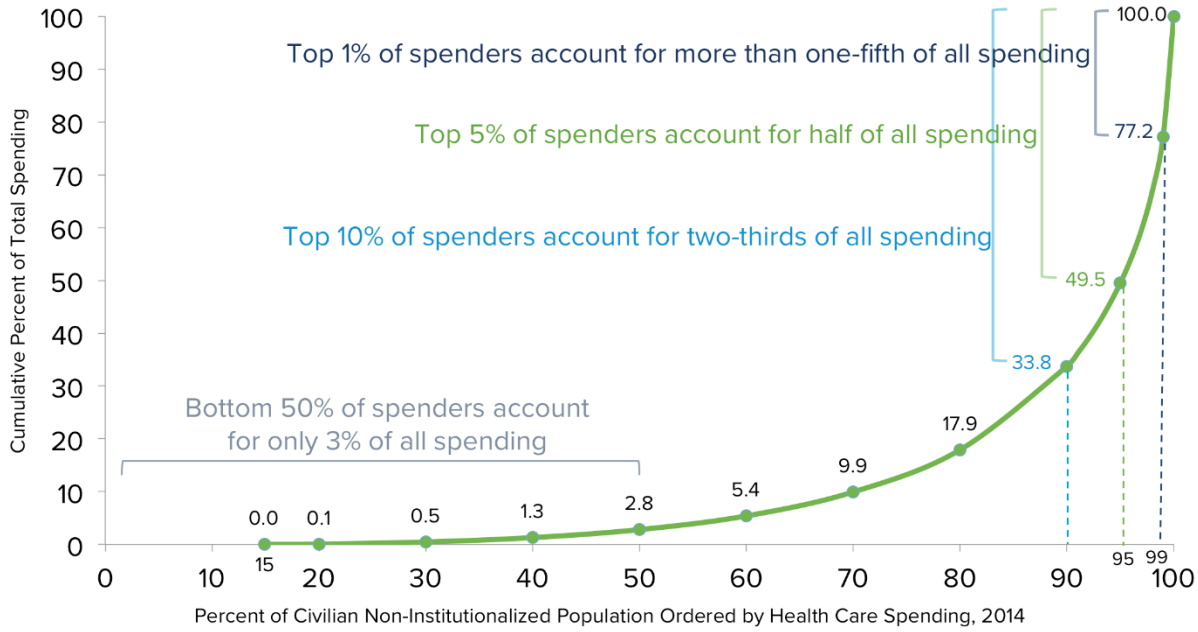


Figure 1: Cumulative total spending against percentage of patient ordered based on healthcare spending plot. (NIHCM Foundation, 2012)

The top 5% of high-cost patients spent an average of 47,000 dollars a year while the top 10% spent an average of 31,000 dollars a year. As you can see in the chart that the total cost of the top spenders significantly out weigh the total cost of the majority of patients. This observation comes from the most complete and comprehensive source of data on cost and use of healthcare and health insurance. MEPS is a "set of large-scale surveys of families and individuals, their medical providers (doctors, hospitals, pharmacies, etc.), and employers across the United States" on which health services are

used and the cost of these services.(Health Human Services, 2016) It is a solid source on which to gauge the overall nature of high-cost patients. Although this data was pooled a few years ago, this trend will nevertheless continue to be prevalent today and years to come if prior trends continue. In a study that aimed to better understand health trends in health care expenditures, the study collected and analyzed various independent sets of expenditure data dating back decades to get coherent image of what high expenditures timeline would look like. (Shown below in figure 1.2)

Distribution Of Health Expenditures For The U.S. Population, By Magnitude Of Expenditures, Selected Years 1928–1996

Percent of U.S. population ranked by expenditures	1928	1963	1970	1977	1980	1987 charges	1987 payments	1996 payments
Top 1 percent	–	17%	26%	27%	29%	30%	28%	27%
Top 2 percent	–	–	35	38	39	41	39	38
Top 5 percent	52%	43	50	55	55	58	56	55
Top 10 percent	–	59	66	70	70	72	70	69
Top 30 percent	93	–	88	90	90	91	90	90
Top 50 percent	–	95	96	97	96	97	97	97

Figure 1.2: Table of distribution of health expenditures for the U.S population. (Berk & Monheit, 1992)

As you can see in the table above, the truth is that the top % of patients have consistently accounted for the majority of costs in the US healthcare system. Even going back to the 1970's, the top 5% of population was spent 50% of total costs and the top 10% spent two thirds of the total cost.(Berk & Monheit, 1992) This is a glaring time-worn problem of the US healthcare system. Furthermore, the trend of high-cost patients accounting for large portions of healthcare cost isn't specific to the US. This weighted

distribution in health care spending has been shown in many health care systems all over the world. Both in Australia and China, there have been studies aimed at understanding their own high-cost patients because it is known that high-cost patients have a disproportionate heavier weight in total cost expenditures.(Calver et al., 2006; Miao et al., 2017)

Characterizing high-cost patients?

Going beyond this simple observation, MEPS and other studies have allowed us to further understand what some general demographics of these patients exhibit. Common trends include high-cost patients being more elderly, having multiple chronic diseases, or having more visits and hospital stays.(Hayes et al., 2016) But these are just generalizations off of retrospective population data. At best we can make healthcare management programs and policies that target these patient demographics.(Hong, Siegel, & Ferris, n.d.) This aspect of population predictions goes into more of healthcare utilization solving. Yet this still hasn't alleviated the issue of high costs being covered by a fraction of patients.

What more can we do to further understand high-cost patients? Is there and trend among high-cost patients that tell us more on the dynamics of their high cost? Not all high-cost patients will be the same nor does their high cost reflect uniformity in the incurnment of their cost. Some patients might have a high cost in just one event or a

high cost spread over a length of time/multiple events. Knowing some granularity in how the high cost is further broken down, may give a better idea on what might be needed to better identify them.

The Health Care Transformation Task Force did exactly that. They are a "industry consortium that brings together patients, payers, providers and purchasers to align private and public sector efforts to clear the way for a sweeping transformation of the U.S. health care system". Part of this task force is the High-cost patient Work Group, which "identifies and evaluates key areas that drive costs for patients in health care systems."(The, Care, & Task, 2015) They analyzed literature to make profiles for high-cost patient groups as a way to have a better sense of how patients could benefit from "targeted care management". Instead of having a program, targeting these patient might be a more efficient method. They divided high-cost patients into three sub groups. *Patients with advanced illness*, *patients with persistent high spending*, and *patients with episodic high spending*. These subgroups aren't completely novel as there were many trends describing aspects of what the work group proposed just never in a coherent dialogue/thought. (Hayes et al., 2016) Putting patients in a coherent context allows us to better frame modeling.

Patients with Advanced Illness:

Advanced illness patients are typically those that are at the end of life. These patients given their advanced illness tend to have high cost because the severity of the unique diseases makes it harder to treat. In the 2014 IOM report, 11% of in the top 5% high cost spenders died within one year. Because inpatient costs usually ramp up near death, there could be many ways to improve healthcare costs.

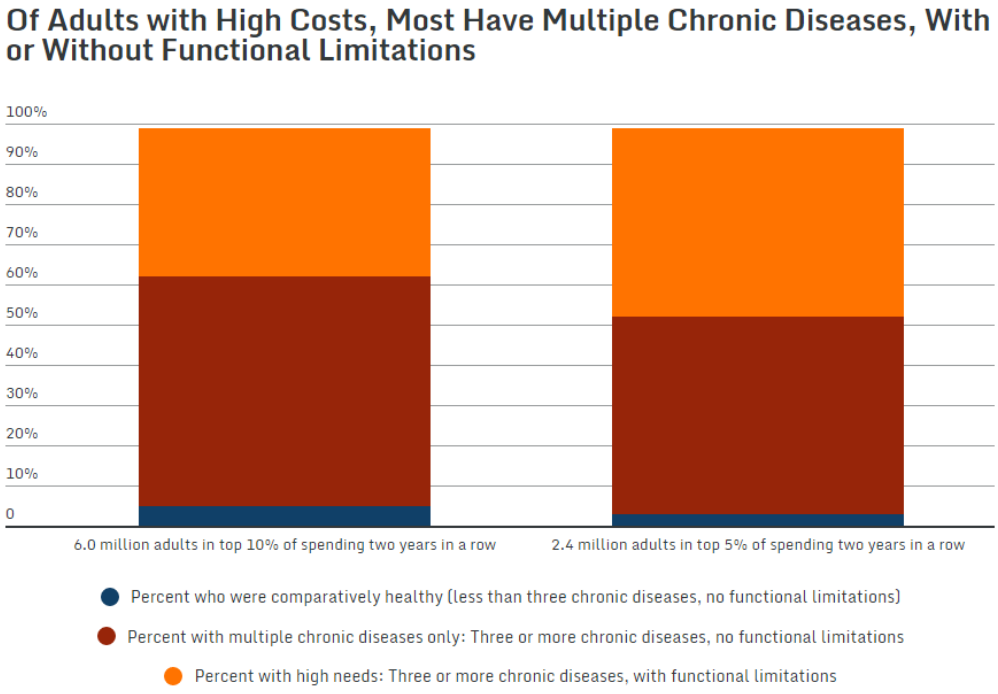


Figure 1.3: Comparison breakdown of high-cost patients with chronic diseases(Hayes et al., 2016)

There is a greater fraction of high need patients in the top 5% than in the top 10% as shown in figure 1.3. Usually some form of palliative care or hospice is the best way of caring for these types of patients. It removes them from the hospital setting which not

only reduces hospital costs, but also gives the patient the most comfort towards end of life. It has been shown that palliative care does reduces cost for end of life care in most places were it is utilized. (Smith, Brick, O'Hara, & Normand, 2014)

Persistent High spenders:

Patients that have multiple chronic condition(s) make up this group. The higher frequency and more time the patient spends at the hospital because of a chronic condition, the higher the expenditure the patients incurs. It is important to note the differences between "common diagnoses from diagnoses that drive spending." "For example, hypertension and hyperlipidemia are widely prevalent conditions, but do not necessarily directly result in high costs, as renal failure, congestive heart failure (CHF), and COPD do." This also applies to patients with cancer and the need for them to get consistent chemotherapy. Patients that have numerous chronic diseases are more likely to have longer hospitals visits and a higher number of inpatient discharges. (Hayes et al., 2016)

Episodic High Spending:

These individual patients are characterized by some sudden event. Following the event, there isn't consistent spending pattern or a reduction in spending. For the most part, these kinds of patients tend to be "seldom foreseeable" given how unpredictable their

sudden event is.(The et al., 2015) Additionally, after the high cost, they seem to have lower levels of spending regardless of intervention. Because it was the sudden event that was the cause of the cost, not the health of the patient. Some example of this are trauma related incidents (i.e Gunshot wound or third-degree burns). Without any hope of intervention, it would be inefficient to target these types of patients with some care management or prevention.

Out of the three sub-groups, patients with advanced illness and persistent spending are the most likely to be predictable and preventable. This makes sense logically as episodic high spending doesn't have any precursor towards occurrence. With a deeper dynamic of high-cost patients laid out, we can ask how do we prevent high-cost patients from getting their expenditure?

Identifying high-cost patients

As reference above all understanding of high-cost patients comes after the fact when the patient has incurred such costs. The question can must be asked:

How are we going to identify these patients who are at high-risk for becoming high cost or are already high cost?

Modeling is currently the only answer towards identification. Modeling takes in data and makes an analysis of the data to be later applied to some dataset for some outcome. Research in modeling high-cost patients has been occurring for many decades.

Yet creating a model is no small feat as shown by the years of research towards this domain. Models use numerous types of data in order to make inferences and understanding of the data. Many make full use features such as simple demographic variables(age, race, gender, and weight) to more medical features(ICD code, the frequency of diagnosis, diagnosis, medications, etc) to even patients expenditure data of a set amount of years(inpatient or outpatient costs). Beyond features, another aspect that needs to be considered is the size of the data set, which may come from a single hospital to that of a network of institutions working together.(Ash et al., 2000; Billings, Dixon, Mijanovich, & Wennberg, 2006)

In one study, a model uses information such as clinical conditions to predict high-cost patients. In their model, they used a logistic regression to identify the top 10% of expenditures for the following year of their cohort.(Lu et al., 2015) They used around 10,000 enrollees with "effect of demographics, count of chronic conditions, the presence of the prevalent chronic conditions, and utilization indicators" as their main variables. The model was valued using the c-statistic. They had a 0.67 as their c-statistic or known as RPC, which is what most models use to show how accurate they are. It is a simple model that just compares a combination of variables for a model. Some limitations were that the dataset didn't have everything on utilization or diagnoses as well as the inherent biases of the population set given it was in an underserved community. This

model is similar to many other ones in that the model is mainly localized in a community of hospital and usually are smaller in number.

Another study compared the ability of different models to predict whether someone will incur high medical expenditures in the second year. In their model, they used a logistic regression to identify top expenditures of a cohort in the second year following tracking of the first year.(Fleishman & Cohen, 2010) They used Diagnostic cost group (DCH) prospective risk scores and a count of chronic conditions and indicators for chronic conditions. Additionally, they tried to see if self-rated health and limitations helped the prediction. As they added progressively added each metric, the prediction improved to around c = 0.836. This is a fairly good value as it measures predictive accuracy of the model. Some limitations of the study were that the dataset was taken from a publicly available dataset which might not be representative of what is seen or available at a hospital. It does at the very least prove that the more data that is added the better the predictive model is.

In the extreme cases, entire provinces like Ontario, Canada are able to pool together patient data from millions of people in order to populate their model. (Chechulin, Nazerian, Rais, & Malikov, 2014) In this study, a model was created from around 10 million patients with 70+ variables each. They used a similar logistic regression in order to utilize the millions of data points. They received a 0.865 as a C-statistic and this is relatively very high. But there are some heavy limitations towards

this kind of model. The primary issue is the size of such a model. It very hard that an institution can obtain data access to millions of patient's records. Furthermore, being able to get a uniform data points across all those patients is difficult. The study even addressed such clear limitations.

Instead of modeling high-cost patients entirely, some models predict an ancillary variable such as hospital readmissions which can be a proxy to cost. These are easier to predict as it more directly with other forms of variables to focus on.(Kansagara et al., 2011) A systematic review of 30 models that predict hospital readmissions shows that most of the current models for clinical purposes perform poorly. These were retested externally for assessment. The most common outcome was 30-day readmissions. Other types of predictions were for risk-adjusted readmission rate comparison, identify high risk patients for early intervention, or hospital discharge identification. Some did, in fact, have good accuracy scores, but only in specific scenarios.

There have been countless models over the years in thousands of different locations with thousands of different patients, yet none of them have been clinically impactful for widespread usage. So, what is wrong the problem with modeling high-cost patients?

The problem with models?

Despite all this work in modeling high-cost patients, there hasn't much success in creating a model to consistently and reliably predict high-cost patients. Of the models referenced above none of them can be used outside of their own hospital setting. At best they provide a proof of concept. What is wrong with models and the current field of predicting high-cost patients? There are plenty of new innovations towards using big data in healthcare. They usually fall under three categories: "More and more data, especially resulting from mobile monitoring; better analytics using new machine learning and other techniques; and meaningful recommendations that focus on prediction, description, and prevention of poor health outcomes"(Ghassemi, Celi, & Stone, n.d.). Now these are all good, but there is caution in the over reliance on big data and modeling. The current field of predicting high-cost patients is saturated with models and the usage of big data in order to solve this. It makes sense though as more and more data does improve the predictive value.(Ghassemi et al., n.d.) As shown in the model with 10 million patients, the model accuracy was higher than the rest. But it isn't probable to have millions of patients readily available for a model to use. A query in the number of publications with key words relating to predicting high-cost patients in the last 70 years shows a huge rise of publications under this topic. Last year had around 33,000 publications with the key words of "predicting high-cost patients".("Predicting high-cost patients - Semantic Scholar," n.d.) Usually usage of

different methods use variations of the same premise. X variable or metric with Y number of patients with Z data set. Every permutation of these generates a unique model that claims to predict high-cost patients to a certain accuracy. Furthermore, many studies just reuse/compare old models as a way of accounting for the limitations of the ones preceding them.(Fleishman & Cohen, 2010) There is plenty of claim to these models as they do generate a degree of accuracy, but many fall short of being clinically useful outside of the setting devised by the model. Many models are limited by size given the institution or by complexity of the model. Most have gone towards the trend of approaching the model in bigger scales, but is the solution just to make a bigger model or a more complex model? Even when a the model used by the healthcare system in Ontario with 10 million patients had a high predictive value they even stated that their limitation was the size and the impractically of such a big size to other locations. If every institution needs to make a model for their particular location, this is impractical. Why should each institution make one? It is redundant and wasteful.

An example of this was the recent trend of hot spotting, which is the "the strategic use of data to reallocate resources to a small subset of high-needs, high-cost patients." At first it was considered as a revolutionary idea to reduce cost and in the beginning, it was claimed to cut a huge percentage of cost, yet as time progressed it was later suggested that the perception of reduced cost were due to overstated results. The trick was perception of the regression to the mean statistic because if the value is

extreme on the initial data point, then it is more likely to be average on subsequent points. Thus patients who might be high cost will be lower cost the next time regardless of intervention. There is even a study that involved 15 randomized trials to test hot spotting/ care management programs which resulted in no difference between enrolled patients and patients receiving such extra treatment targeting.(Greenberg, 2016) This is not to say that such hotspotting programs have failed everywhere. Just there needs to be more analysis on if something works and why it actually works. There was so much initial expectation that it worked regardless of whether the results were validated. Are the actual identification methods actually identifying a common trend or just something that is a trend of that local population?

What needs to be known right now to proceed?

Instead of just making more models and doing something that is slightly different, but still the same, we need to understand what is the data that we actually have and what is happening to these patients. From there we can understand what is happening to the patient. This stems from reliance on the data itself. Models are simply treating data at face value and trying to draw meaningful connections from data without looking at what the data holds. Instead of saying this patient that is high cost exhibits these data points, think more on this patient was high cost because of X. Additionality it is important to talk about the streetlamp effect in, which an example is

that a drunk man goes to a lamp post to look for his keys even if it isn't there because it's the easiest place to look for them. Variables and features of EHR are easy to look at for models because they are so readily available, but this might be the wrong approach. Instead look at the data we need and from there we can generate a more accurate model. A study that was done with machine learning providing the mechanism for prediction cited how useful it would be to partner with the use of expert knowledge as a way to better tailor their model.(Moturu, Johnson, & Liu, 2010)

What can be done?

Now that we have framed the full problem, what other approaches are there? Instead of treating the data as our solution, would putting more value in the data itself help? If many just use variables or metrics to find some level of correlation, we need to understand why there is a correlation or even a lack thereof. We simply know they are high-cost, but no way of knowing what caused that high cost. From there we can understand what is happening to the patient. Instead of saying this high-cost patient exhibits these data points, think more on this patient was high cost because of X. As referenced above, some high-cost patients will always have high cost because it's in the nature of their health regardless if it was prevented. "The goal should not only be to target the highest cost group but also aim interventions at preventing individuals from entering into this group in the first place."(Lu et al., 2015) There is so much attention

placed on quantitative modeling. Quantitative data do provide reliable data as they come from claims data compared to qualitative data that is more from patient reported information. Yet the health care transformation task force that helped outline the types of patients that made up high expenditures suggested applying more qualitative approach towards predictive modeling. Expanding that thought process, I sought to apply qualitative measures towards understanding the dynamics on why current models can't succeed.

Methods

Overview

The study is looking at retrospective clinical notes in order to understand the primary reason behind the high cost of the patient. This is done to then better understand the breakdown of what is incurring the high cost for a group of high-cost patients and is there a way to better categorize where the cost stems from.

Population and Study Sample

The population of the study comprised mainly of patients from the University of Washington Medical Center and Harborview Medical Center, which fit under the University of Washington Medicine umbrella. These patients have the common denomination that they fit the criteria of having health care expenditures in the fiscal

year of 2016 that puts them in the top 10% of expenditures. Furthermore another subgroup of patients were chosen from Northwest Hospital which is a community hospital within UW Medicine. They also fit the same criteria of having the top 10% of all costs within that institution.

Sample Size and Selection of Sample

The sample size consisted of 100 patients from Harborview and University of Washington Medical Center. Another 100 patients were taken from Northwest hospital. These patients were taken off of the clinical data repository. EPIC was used to also query through the EHR of Northwest Hospital. From there, a random number generator chooses 100 patients out of each list. In order to match the visit date entry from the query to the entry in the EHR platform, the patient's MRN was noted and verified, but thrown away after the study.

Patient List:

The list of high-cost patients came from fiscal data coming from the 2016 year for the UWMC/Harborivew patients and the fiscal data coming from the 2017 year for the Northwest hospital patients. Both lists contained MRN numbers of the patients.

UWMC: The fiscal data contained the breakdown of patients cost for any visit during the 2016 year. This helped with identifying what was the exact visit the high cost was

billed on. Also this helped in determining the number of visits and their time between each other.

Northwest Hospital: The fiscal data included the inpatient and outpatient costs for both indirect and direct costs. This helped bridge the costs in the fiscal data from the query and particular visit entries in the database. I did not use indirect costs as it many refers to costs not relevant to the actual service towards the patient (fees for floor maintenance, administration fees, etc).("Components of Indirect Costs," n.d.) I also did not use outpatient costs because they would not be loaded onto the EHR database. An issue that I identified was that there wasn't an easy transition between the billing list of patients and the database. The billing list only gave financial case identifiers numbers were aren't recognized readily in the interface. Instead, I resorted towards querying a list of all visits for 100 patients to show both the HAR numbers and the ICD/diagnosis ranked.

From there, the source of clinical data came from the UWMC database through the ORCA interface. The ORCA portal was accessed remotely through a Citrix portal.

Security of Data:

Getting data access for both sites involved going through the institutional review board approval process. The study went through expedited review given its lack of significant risks to the patients and accidental findings towards patients. This was all that was needed for UWMC/Harborivew access, but for Northwest Hospital an added

step was needed. Another approval process was added given that Northwest Hospital was outside the immediate system of the UWMC umbrella. With access to both system , there was numerous EHR platforms access given, but the only ones that used included ORCA for UWMC and Cerner for NHW hospital.

Interface - Data collection:

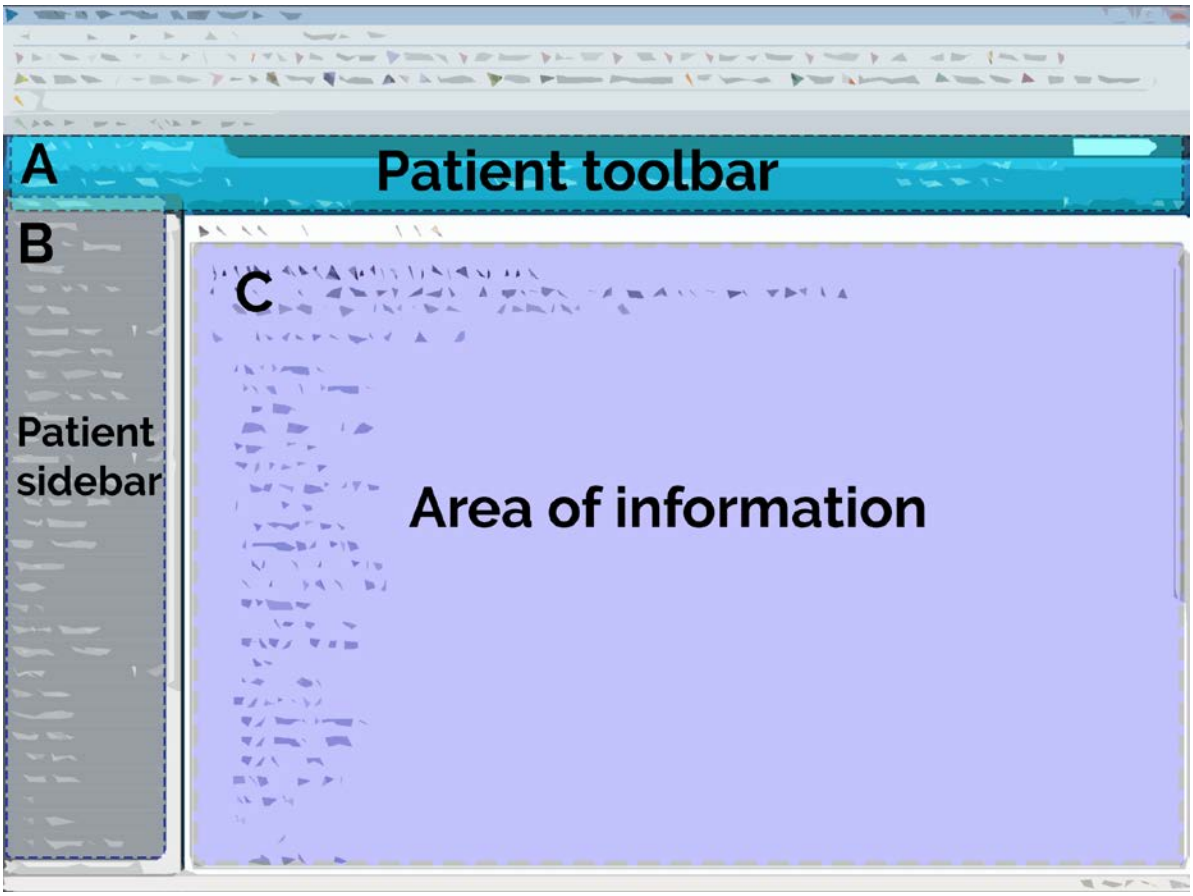


Figure 2.1: Shows the complete interface of the portal. Areas of importance are mainly the A. Patient Toolbar, B. Patient Sidebar, and C. Area of information.

Section A is where all of the general shortcuts for all patient data, but in particular MRN search function is the most important. In section B, this contains all of the patient data

sections. The main sections that I use include the clinical notes and EpicCare within the sub category.

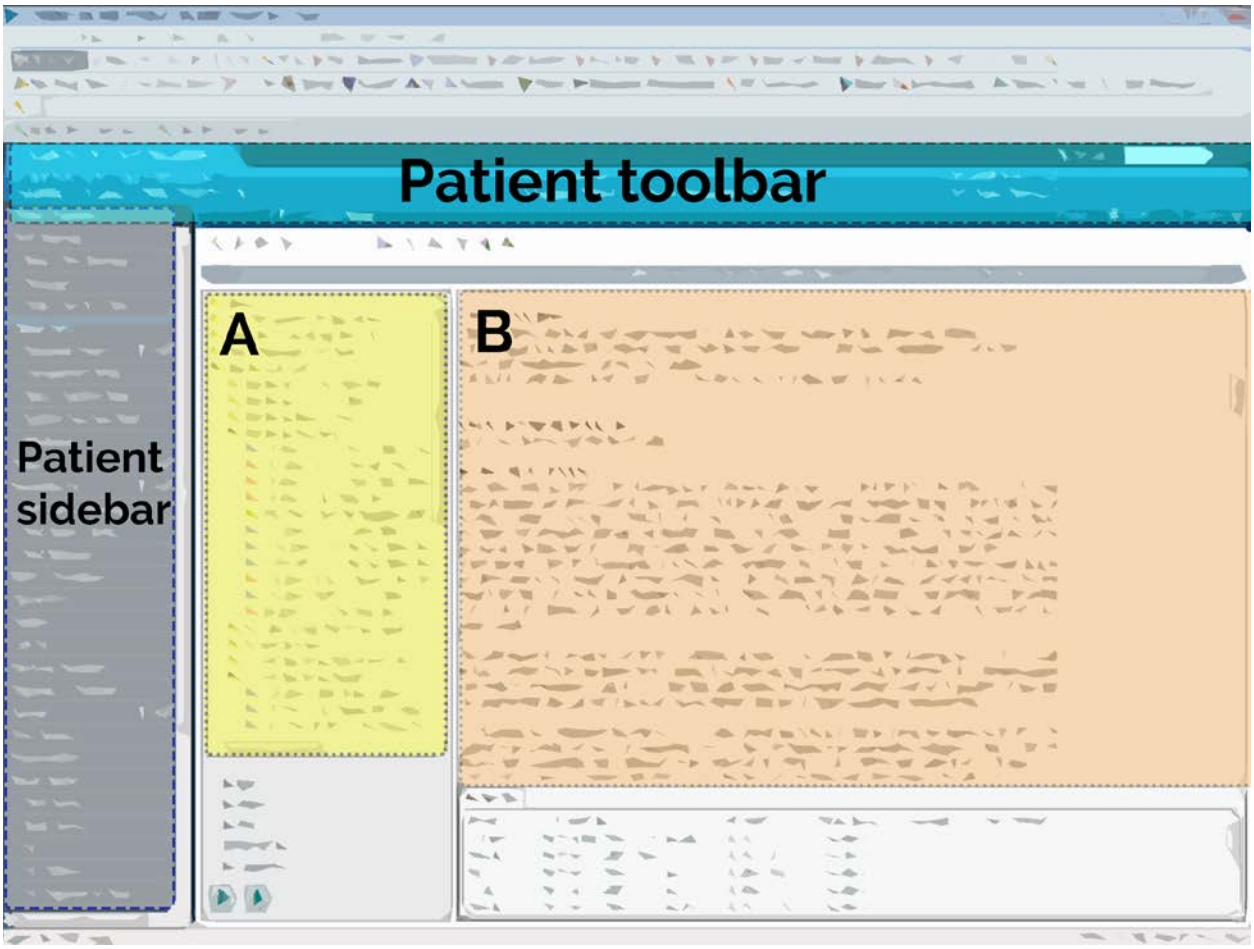


Figure 2.2: Shows an example of what a patient's clinical note would look.

Section A depicts the listing of visit notes based on date. This can be further edited for type or by physician. Section B contains the actual notes. Note that this does change based on what type of note is shown. Procedure, discharge, and summary notes all alter what is shown and the format of each.

Interface - NW Hospital:

For the Northwest Hospital data, Cerner had to be used to access patient data. They were not initially part of the UWMC system and thus maintain a different system.

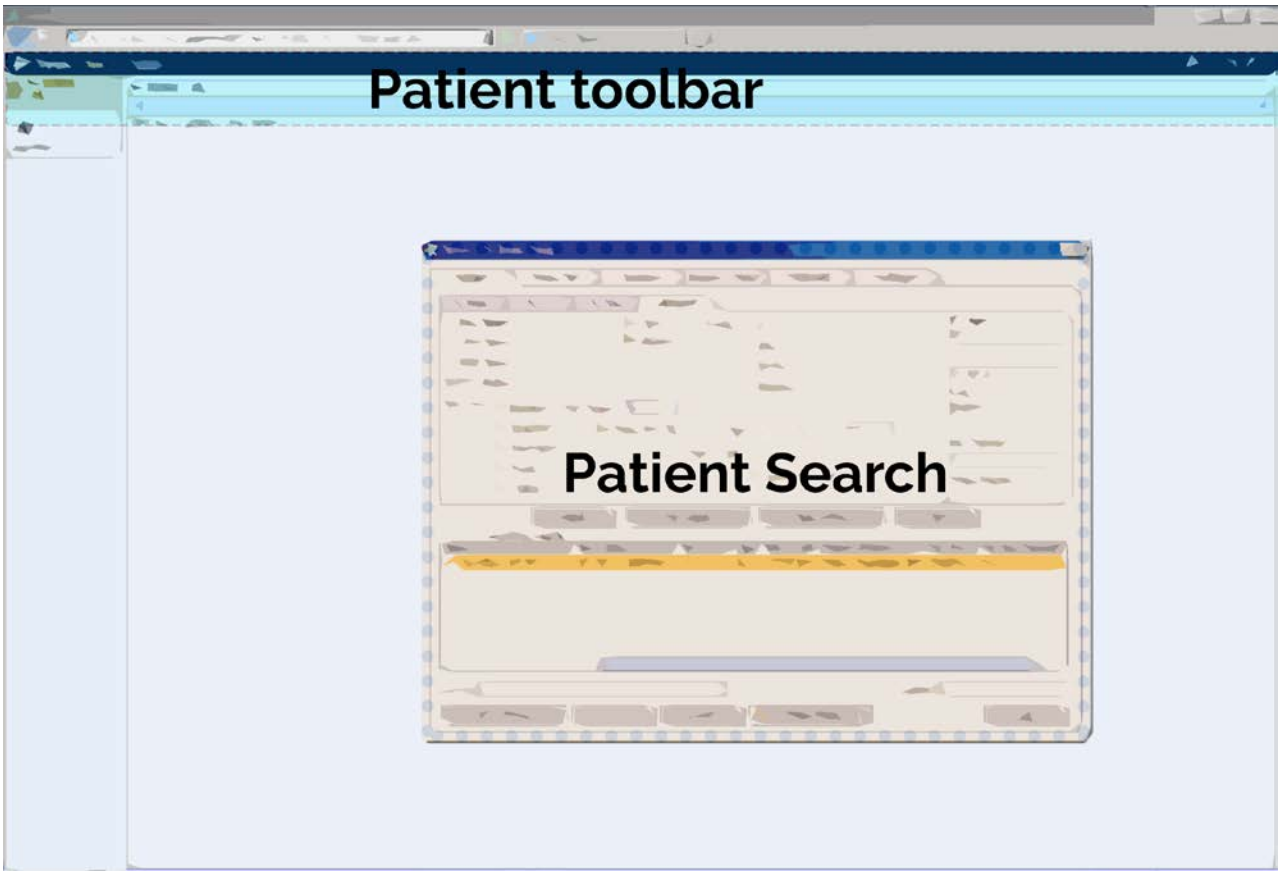


Figure 2.3: Display of search window in Cerner.

It allows the easy of search based on multiple types of ID's (most notably MRN and account number), but not HAR which was the billing list of all patients had.

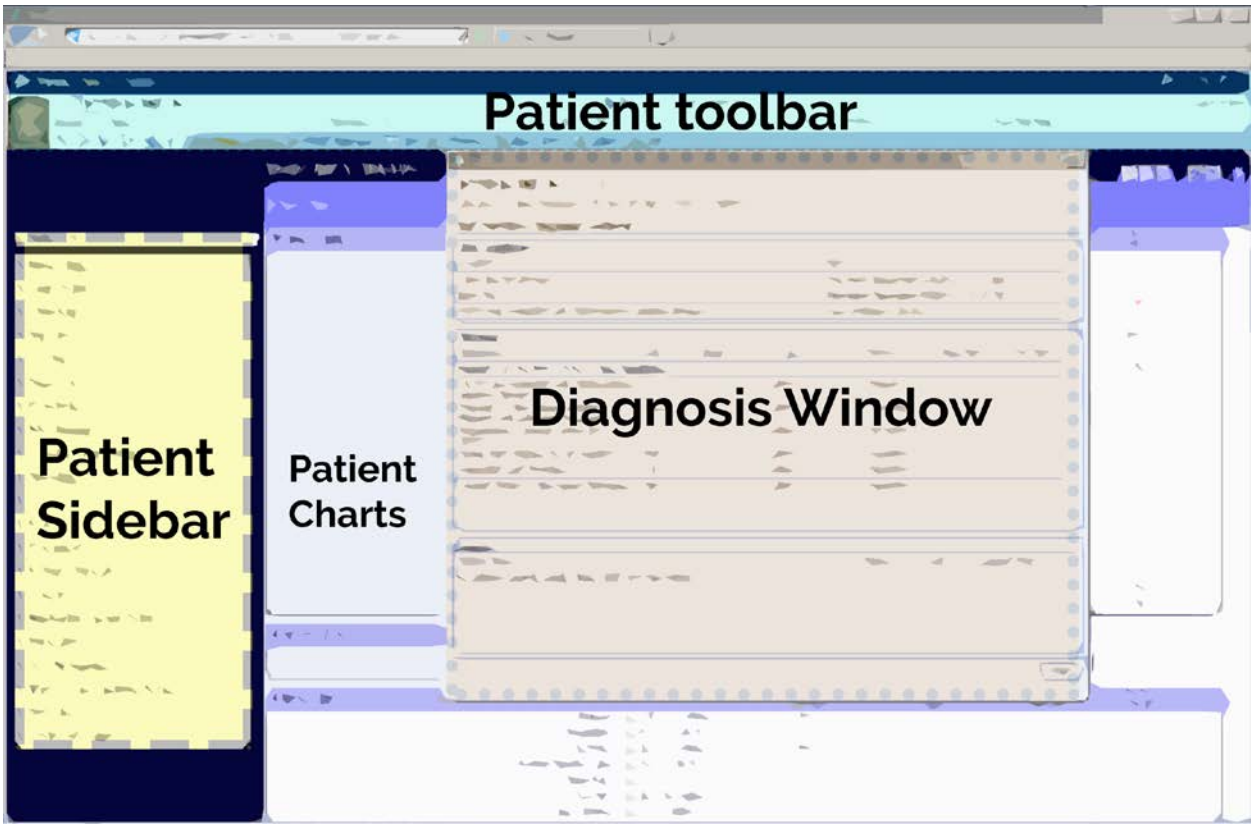


Figure 2.4: Display of a patient chart in Cerner.

The left panel shares similarity with OCRA's patient sidebar except with less options. The current window showcases a list of diagnoses that the patient is coded with. The primary one being chosen and listed among the remaining secondary diagnosis . Beneath the diagnosis window are the patient charts.

Collection of Data:

UWMC:

Given how much excess data was available through the EHR interface, there needed to be a degree of limitation on what was gathered. Starting with what other models have

done, I started with most of the common variables. I ended up with a general template of demographics (age, sex, etc) then number of visits in that year with time of those visits, medication, etc. Yet after comparing what I had initially with what was in the EHR, I realized that my template wasn't going to achieve anything because my template didn't tell why the patient was even high cost. Starting over, I instead started with the EHR/cost and worked backwards. Looking at why the patient was admitted and finding the source of the high cost in the same mindset if a physician had to look through the EHR to how best predict if a patient would later become high cost. Starting with the cost data, I would figure out what was the cost dynamic of the patient. If the cost was divided equally over a few days, then I would try to understand if that was coincidence or based on some underlying cause (such as a constant medium cost over 4-5 visits vs a large cost on 1 event). I began at the clinical notes. In trying to find a snapshot on what happened with the patient. This took iterations on different clinical notes. In clinical notes, I looked first at the discharge summary which contains listed summary of the whole visit. This would usually tell me exactly what happened. But sometimes the discharge summary could be vague if written by different physicians or might not really tell much about the whole incident. So, if the discharge summary didn't have much information, then I moved onto procedure summaries and admit summaries and working my way from there. Procedures do have summaries on what was the reason they were admitted, and the process written up to that procedure. They

also provide the needed information on if there were any complications that occurred during the operation. Once I have a list of all the diagnosis's or the timeline of the patient, I tried and captured the reason of expenditures for each visit. Then I categorized those based on their commonality.

After a few iterations, I stuck with this layout and choices for what information I need from patients notes for UWMC/Harborview.

Table 1: Listing of all types of data collected

Simple Demographics			
Age		Deceased Status	
Patient history			
Record prior to 2016		Past Medical History	
Visit info			
Visit Date(s)	Visit Cost	Number of Visits	Total Cost 2016
Visit secondary info			
Transfer indication	Transfer Reason	Procedure(s)	Type of Primary admit reason
Predictability			

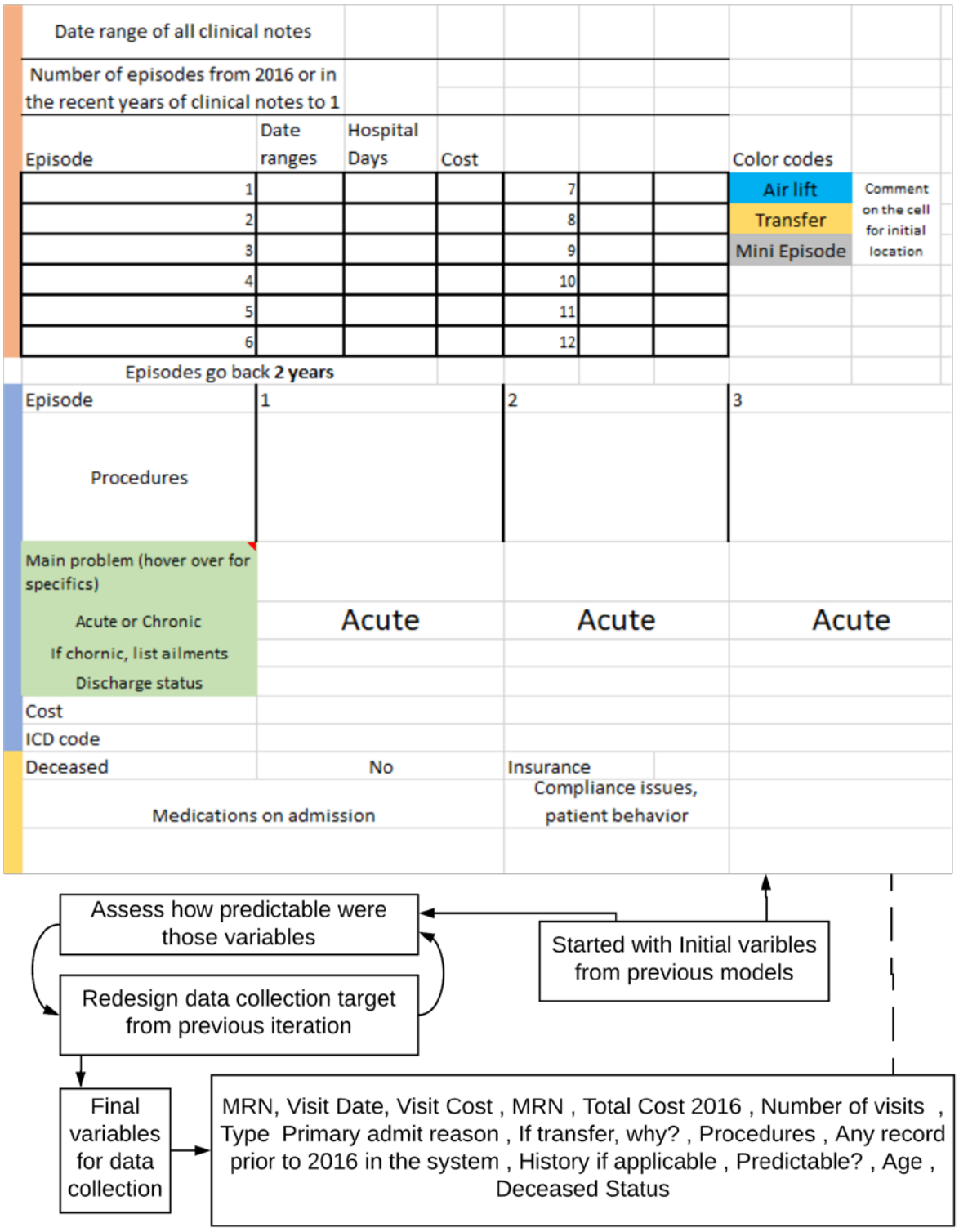


Figure 2.4: A workflow on how the process towards picking the variables.

Northwest Hospital:

Northwest hospital had an entirely different system. The methodology was similar with the workflow in finding what was the reason for the high cost. With the cerner platform, there was a way to look at a diagnosis list that also had primary and secondary diagnosis listed. This provided a very helpful snapshot on what the reasons were. This was crucial in that the platform for cerner was not the most receptive to use and search through. In using the diagnosis list did come with minor inconvenience in that there is no distinction between anything beyond the second diagnosis. A fast workaround was querying through the diagnosis which allowed me to see all the diagnosis rankings.

External Verification

My data collection was later verified on the data collection and analysis by an external physician. The physician would verify blindly and would try their best to also predict the cause of the high cost on a subset of patients that were chosen randomly.

Vignettes:

Process on how different patients were characterized:

Acute (Gunshot Wound): An acute patient and in particular a gunshot patient would typically be recognized first by the lack of medical records before the event date. This just means that the patient hasn't even come to the hospital before. Then through the discharge summary the mention of the gunshot incident or any other acute trauma would be noted. There might be

some confusion if the patient's health condition worsens. Then the end diagnosis might not be the acute accident.

Transfer: A transfer patient is characterized simply by the mention of the reason for admit being related to a transfer from a different institution. Transfers and acute patients both are in common that they don't usually have medical records of the patient available before the event date of the high cost.

Cancer: Most distinctly characterized by the repeat consistent high cost over a set amount of time. This is due to the cost of chemotherapy appointments. Further inspection of the clinical notes would confirm cancer.

Cardiac: No reliable set of trends to characterize cardiac patients. The only way is to simply go into and look at the clinical notes for an indication of a cardiac procedure or diagnosis.

Transplant: Characterized by a cluster of appointments with some singular event preceding it. This usually refers to the referral of a physician for UWMC's transplant services.

Northwest Hospital Additions:

Infection: Characterized by patients with a priority diagnosis of any form of sepsis or infection. Usually is also accompanied by some subsequent infection diagnosis.

Data Management

As with clinical notes, there are issues with missing data and pinpointing what was the root cause. Many times when looking at the clinical notes, there can be many overlapping conditions and diagnosis, furthermore a patient's history could go on for years so trying to be confident in what the root problem was involved a lot of tracing

and confirmation with an external party. Patient data were stored in password protected documents with MRN's only recorded if necessary for data access in the portal.

Ethics and Human Subjects Issues

The study passed and IRB study approval. The terms of the approval were expedited given that there was limited risk towards patient's data.

Results

UWMC/Harborview Patient Age:

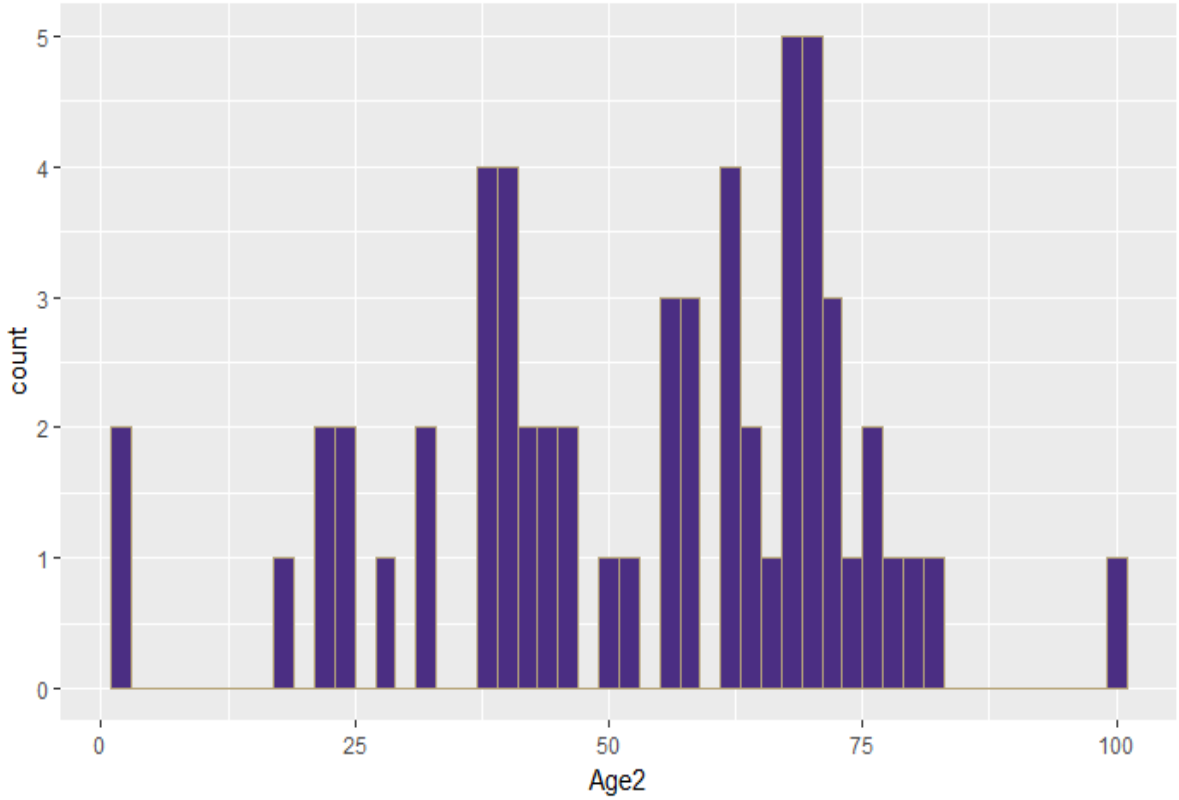


Figure 3.1 Histogram of UWMC/Harborview patients

Northwest hospital Patient Age:

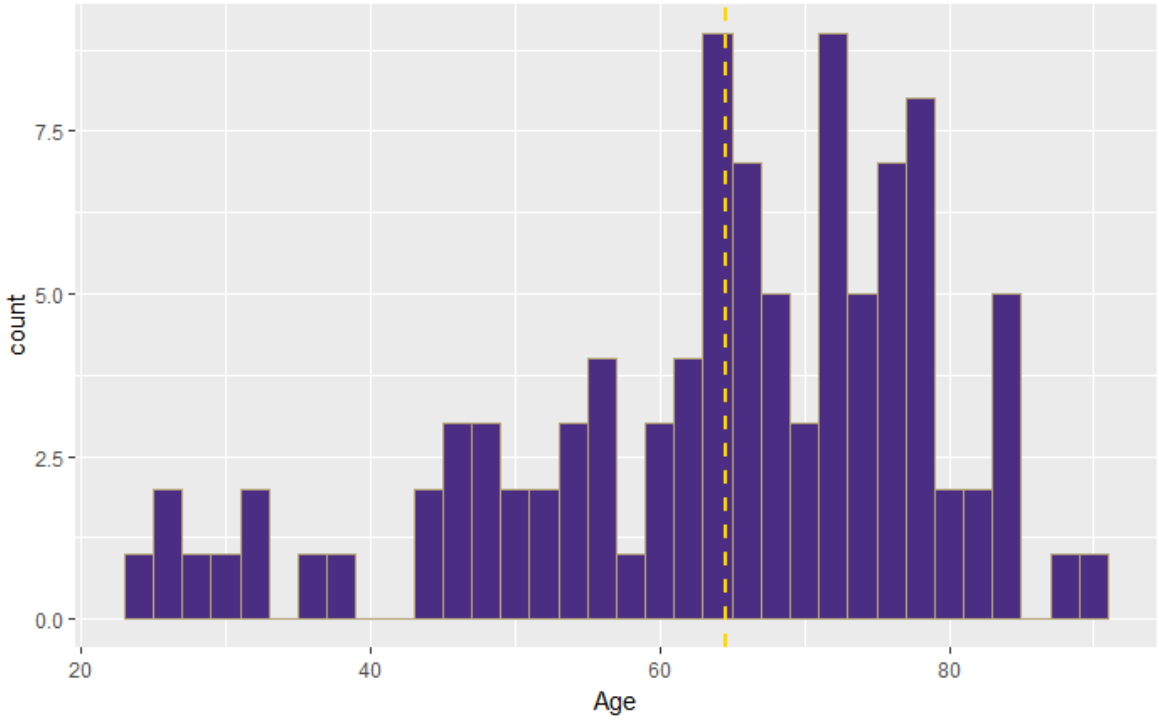


Figure 3.2: Histogram of NW Hospital patients

Statistics:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
24	56	66.5	64.53	76	90

As shown in both histograms, the age of high-cost patients is more weighted towards the elderly. This is consistent in the overall trend that high-cost patients are typically older. Compared to most hospitals have patients that have a relative similar patient size for the 18-60 year old age group. (Weiss & Elixhauser, 2012)

Visit breakdown:

UWMC/Harborview accounts for most of the acute incidents in King County as well as severe injuries in the WWAMI system. Northwest Hospital is more oriented to inpatient and outpatient care, which serves as a community hospital.

This difference can be seen in the dynamics of how long each patient's visits are.

UWMC/Harborivew Data:

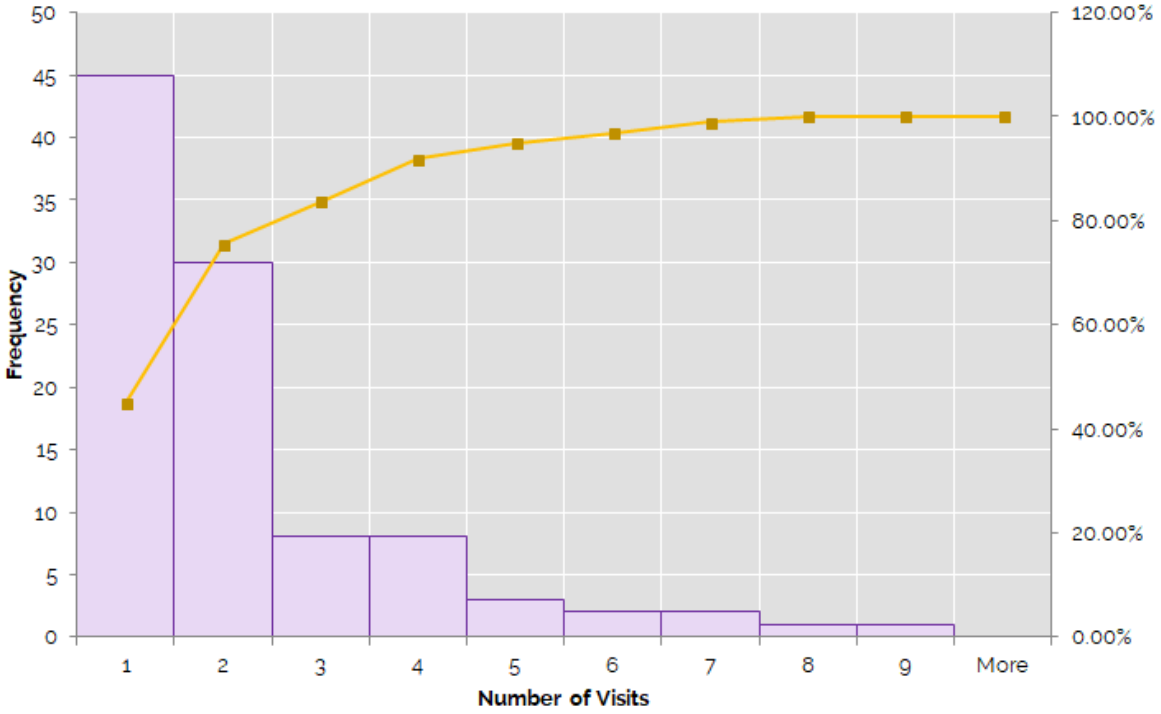


Figure 3.3 Frequency of the number of visits per patient for UWMC/Harborview

The graph above shows the breakdown of how many visits each patient in UWMC/Harborivew had. As you can see most of visits are weighted towards the fewer number of visits with only around 20% being more than 3 visits.

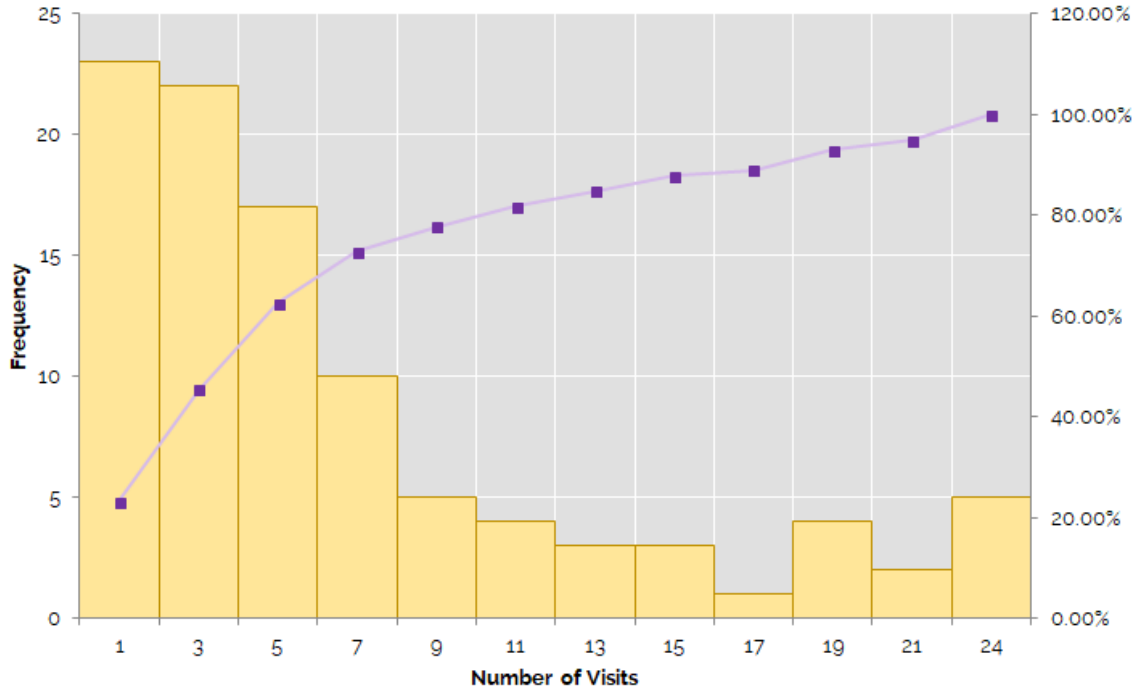


Figure 3.5 Frequency of the visits per patient for NW Hospital

Compared to UWMC, the visit distribution is different. Instead of it being more weighted towards few visits, they are more spread out comparatively.

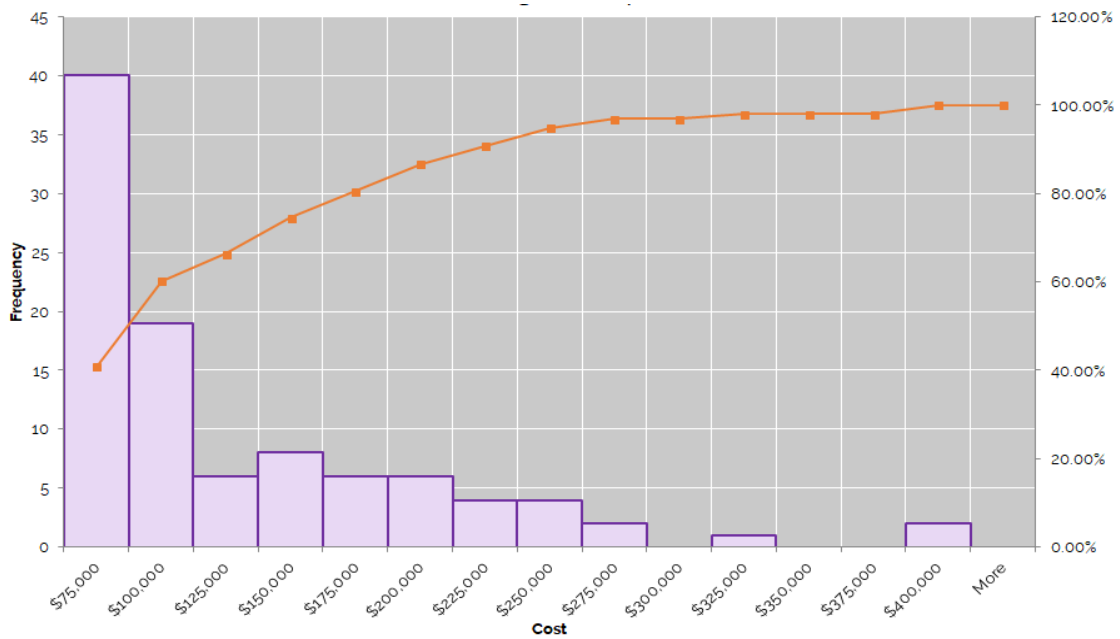


Figure 3.6 Costs of high patients at UWMC/Harborview (Histgram)

The graph above shows the breakdown of patients costs the 2016 year. Again with a similar trend to the visit frequency. There is a weight towards numerous lower end high-cost patients. There is a slight linear trend in cumulative costs, so although subsequent visits aren't as numerous they still impact total costs just as equally.

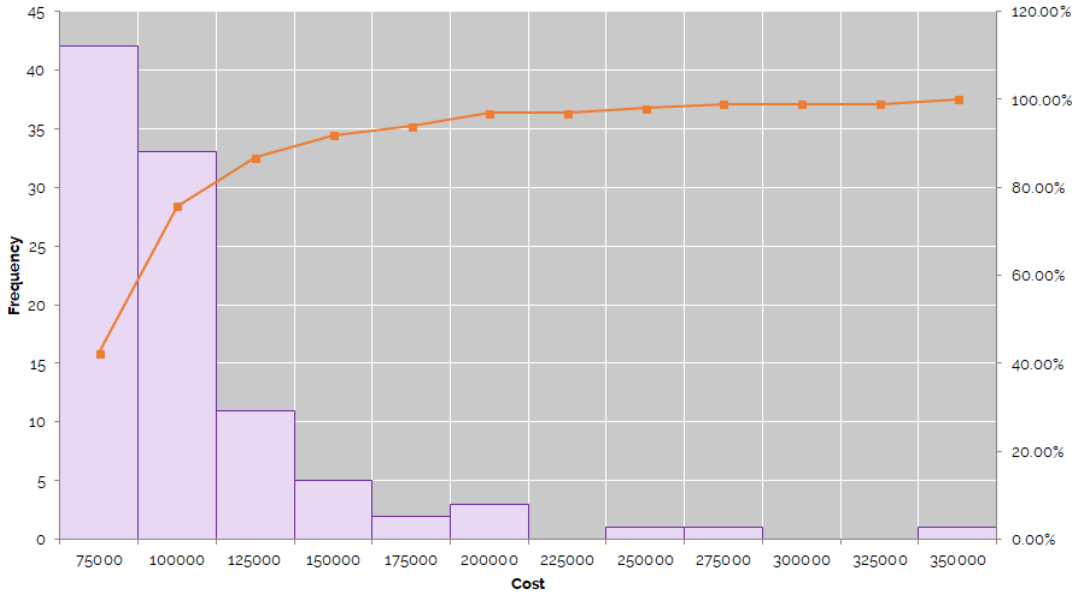


Figure 3.7 Costs of high-cost patients at NW Hospital

Compared to the cost breakdown from Northwest Hospital, the costs are slightly more weighted towards the lower end with 80% of patients under 100,000 vs 80% under 175,000.

Breakdown of visit dynamic at UWMC/Harborview:

1 Visit:

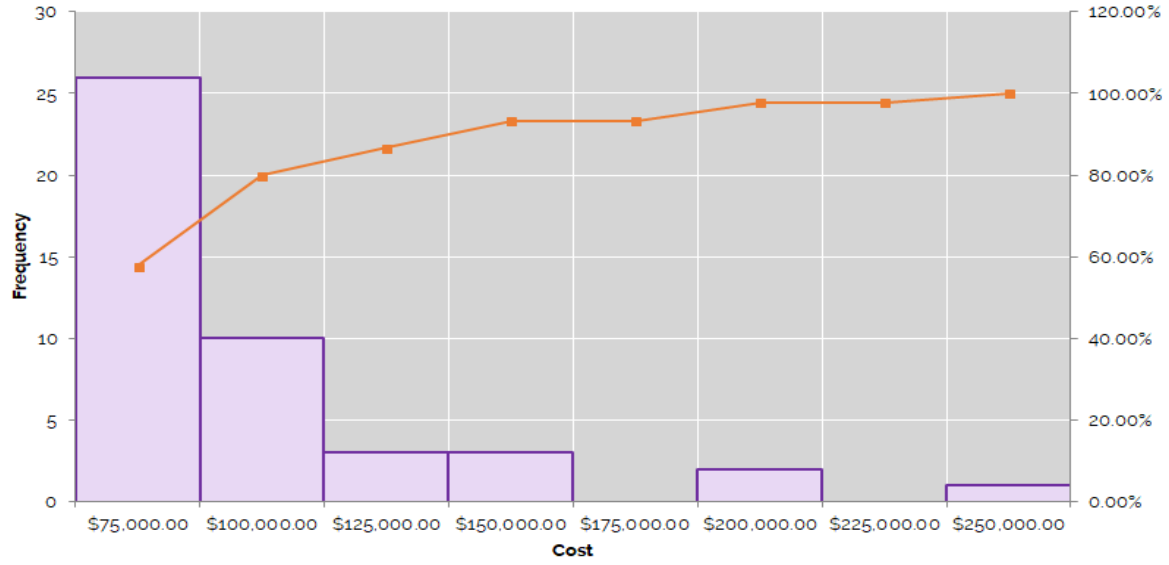


Figure 3.8 Chart of cost for 1 visit

Patients that are typically have only one visit that are high cost are mainly that of a lower bond cost in the full range of costs. Around 70% of patients make up one of the lower bin cost categories.

2 visits:

With two visits, costs were varied between the first and second varied among the respective cateogry. This can be contributed to the inherent dynamics of the categories. The exact ranges can't be used statistically given the low number of patients that did have 2 visits, but the relative positions can be used for understanding.

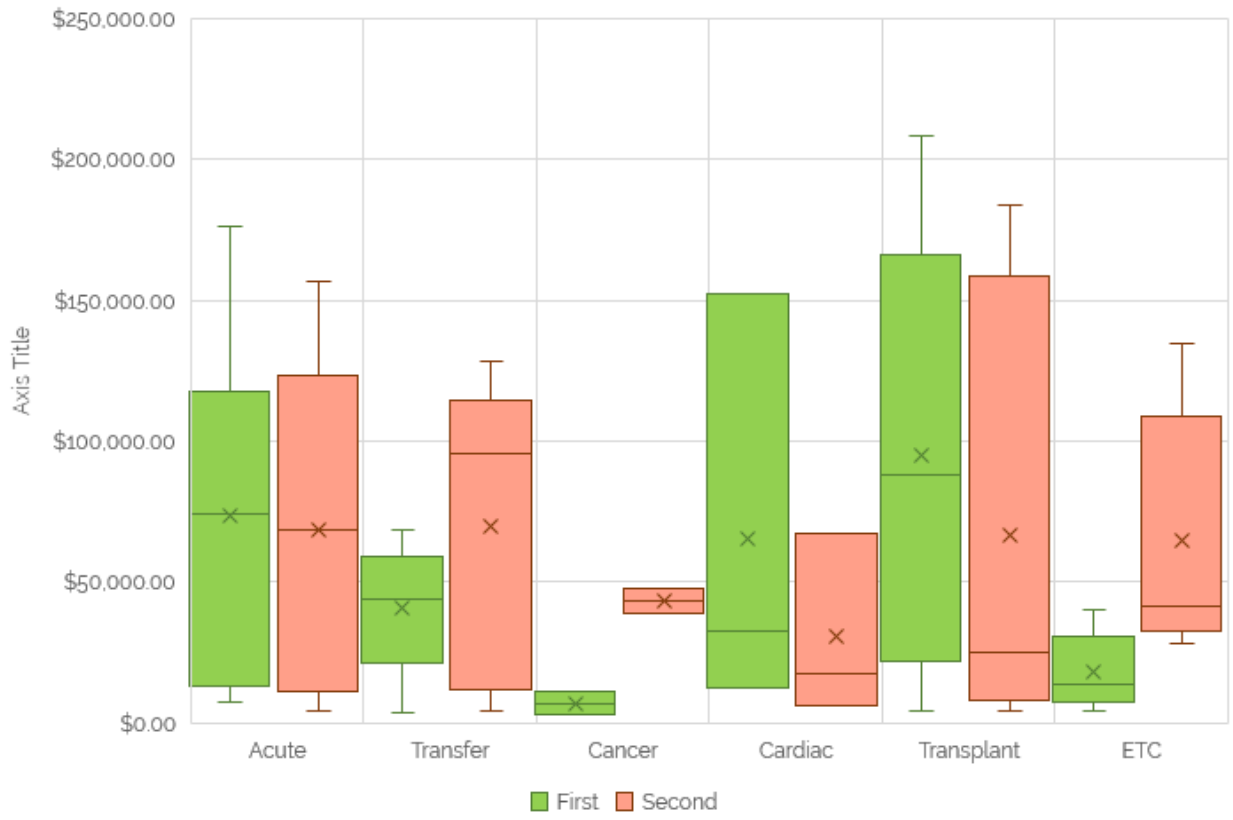


Figure 3.9 Comparison of costs for 1st day and 2nd day

Clinical Notes Data:

From each patient's clinical notes, I sorted out whether a patient's high cost was predictable or not. It can be seen that there is at least a two fold difference between patients that can't be predicted from those that can. Out of the 100 patients at UWMC, 26 were predictable and 71 weren't, and 3 patients it was impossible to determine. Out of the 100 patients at Northwest Hospital, 26 were predictable and 64 weren't with 9 patients that weren't possible to determine reliably. [Figure 3.10]

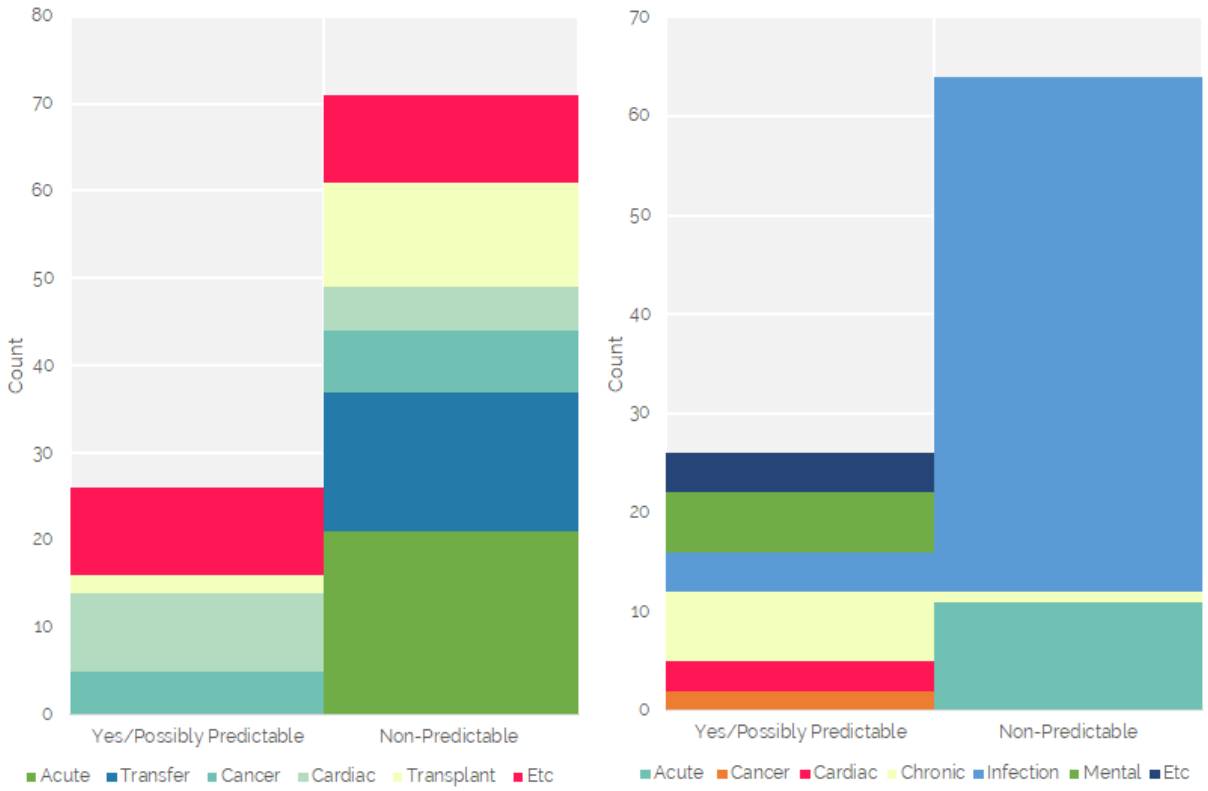


Figure 3.10 Comparison of UWMC/Harborview & NW Hospital tally of predictability

Referring to UWMC/Harborview, a sizeable chunk of the non predictable patients stem from transplant, acute, and transfer patients. Transfer patients typically will be transferred from a different hospital that is outside of the target hospital network. Thus any medical history or information from the original hospital isn't able to be shared in the target hospital system. It is impossible to actually predict patients that transfer since they don't even exist in the system. There may still be a record of the patient after the fact through a third party medium through the platform, but the information does not exist in the target hospital. Acute patients are similar in that there really isn't any reason to predict these. They happen episodically. No form of information in EHR can hope to

predict these patients. Finally, with transplant patients, because most of them are getting the transplant at the hospital like UWMC, but the preparation isn't really located at UWMC. They are referred given they have the facilities to carry out the operation. From this, it is less of the patient being predictable than it is just simply an inherent high cost.

After classifying patients under their respective predictiveness, I further categorized each patient's cause of their high cost.

For UWMC/Harborview, I broke down categories as:

- Acute: Gunshot wound, motor vehicle accident, burn
- Transfer: Transfered or airlifted from a different hospital
- Cancer: chemotherapy, radiation therapy
- Transplant: heart, kidney, liver, etc
- Cardiac: LVAD, value replacement
- Etc: Covers anything that isn't under the umbrella of the other categories

For Northwest Hospital, I broke down categories as:

- Acute: hemorrhage, acute injury
- Cancer: Chemotherapy, Neoplasm
- Cardiac: Heart Failure, heart implants
- Chronic: Alzheimer's disease
- Infection: Sepsis, multiple forms of infection
- Mental: Paranoid schizophrenia
- Etc: Covers anything that isn't under the umbrella of the other categories

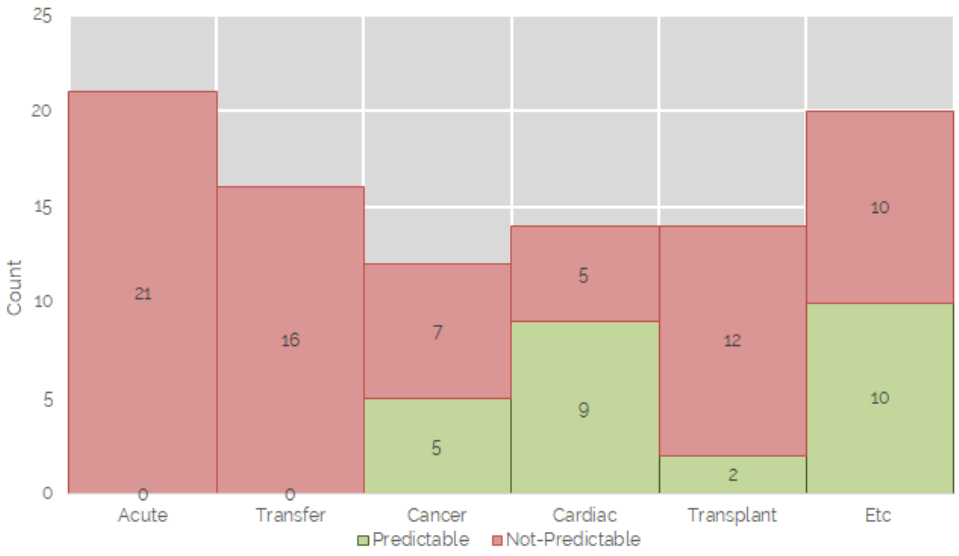


Figure 3.11 Ratio of unpredictable to predictable patients in categories at UWMC/Harborview

Looking at the chart above, it can be seen that acute and transfer patients are clearly dominated with non-predictable patients. Every other category is less one sided with varying degrees of predictable patients.

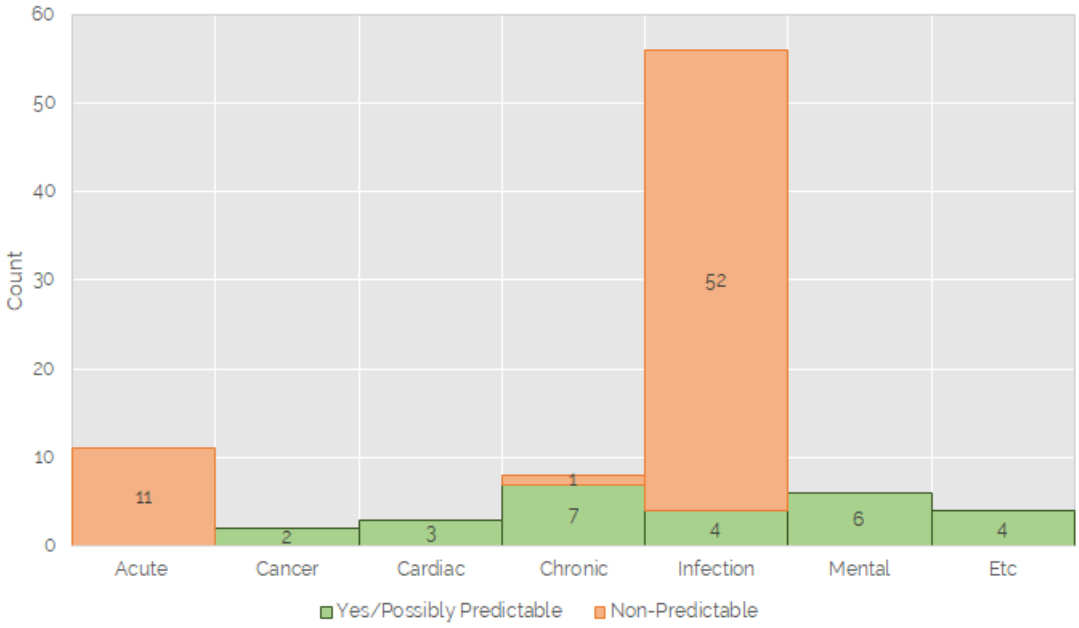
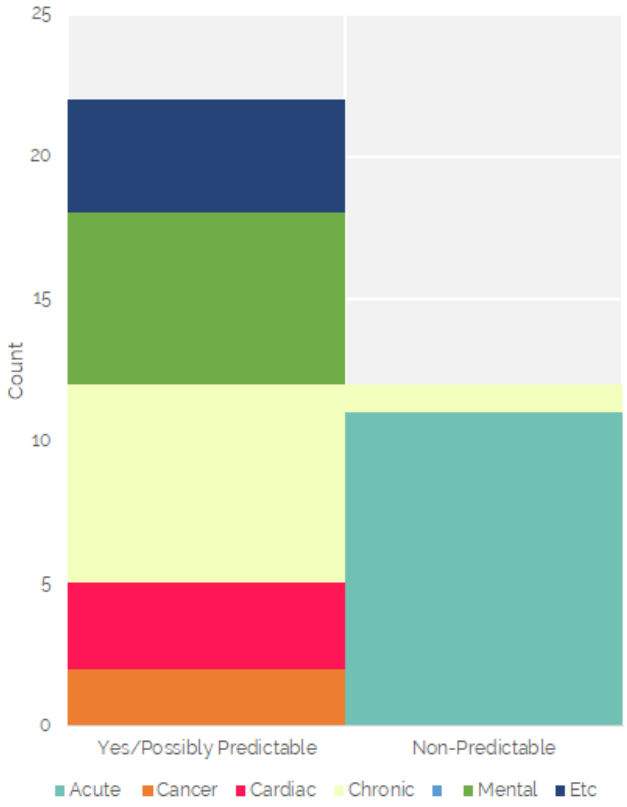


Figure 3.11 Ratio of unpredictable to predictable patients in categories at NW Hospital

There is a significant difference between the two hospitals. Compared to UWMC/Harborivew, there is a difference that more entire sections are predictable. Yet when looking at the impact that infection category impacts the weighting, by taking infection out it changes the dynamic significantly. With infections out of the frame, the predictably becomes greater than that of non predictable patients.



Cost breakdown:

After looking at how the actual predictability of patients broke down, I moved onto the cost breakdown. It seemed important to better understand which costs were greater than others.

UWMC/Harborivew vs NW Hospital:

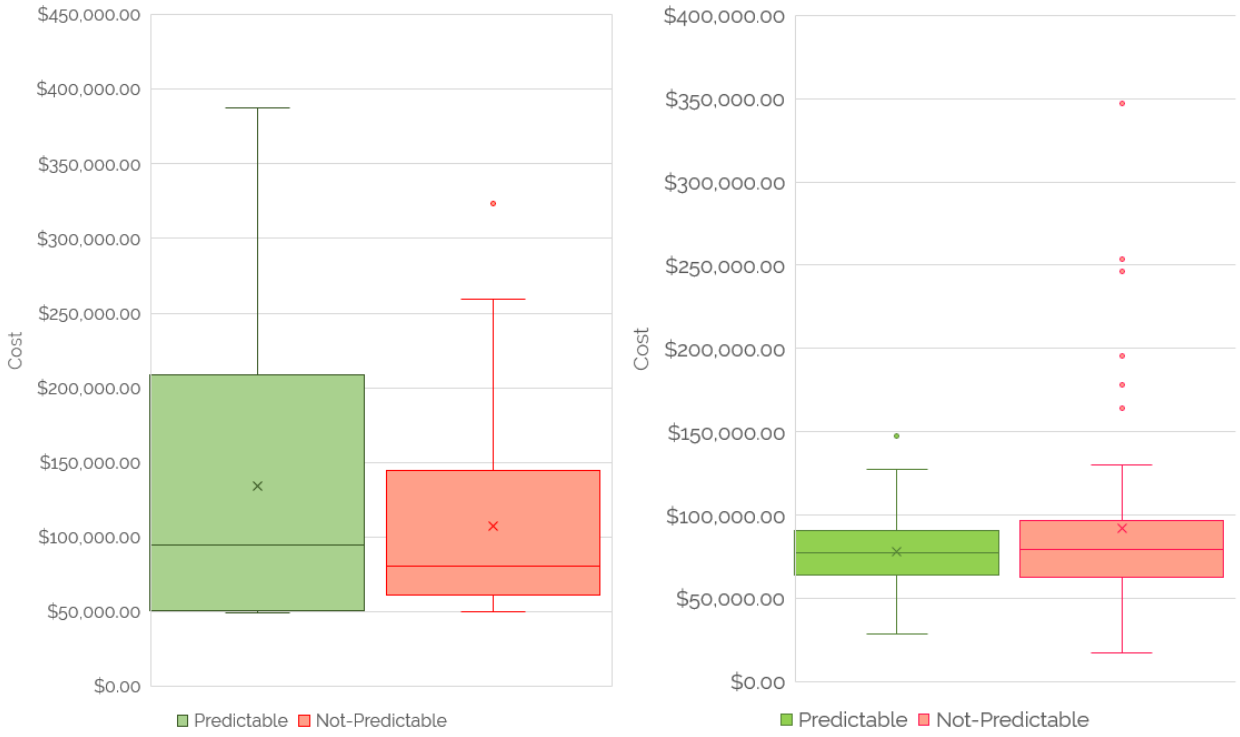


Figure 3.13 Comparison of costs between predictably among hospitals Left: UWMC/Harborivew Right: NW Hospital.

Taking a whole view of the costs with predictably, it can be seen that the high cost of patients that can't be predicted is significantly smaller than that of patients that can be predicted. There is more variability in the cost range of UWMC/Harborview patients than that of the non predictable patients. As for Northwest Hospital, this seems to be completely different with regards to the cost range. Seemingly more tight of a cost fit among NW Hospital patients.

Category Cost Breakdown:

Given that patients preliminary have a high cost that can be characterized by some sort of general health motif, can we see if this can be explained through cost categorization through the sub groups?

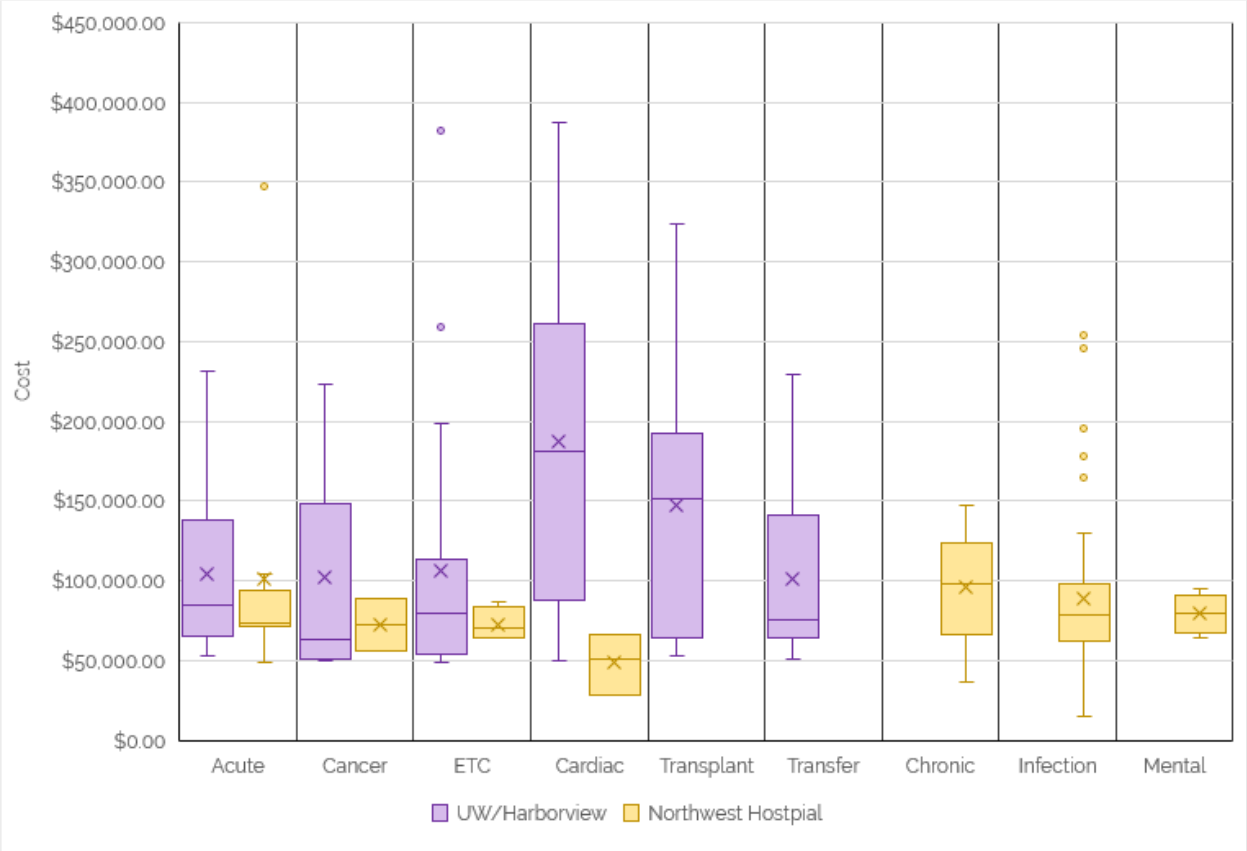


Figure 3.14 Comparisons of costs among categories between hospitals

All categories hovered around a similar cost range, except for cardiac and transplant. Cardiac is seemingly greater in expense among the two. This only happens for the UW/Haborview patients. When looking at the NW hospital data, the cost ranges are a lot tighter compared to UW/Harborview costs. Of the categories that are shared between both hospitals, NW hospital costs are usually less costly comparatively.

Going further to look at how the costs are broken down with predictiveness, comparing the costs of categories that have both predictive and non-predictive patients it can be seen in the chart below. For cardiac, transplant and ETC sections, predictable costs are usually higher than that of not predictable costs while cancer it was the opposite.

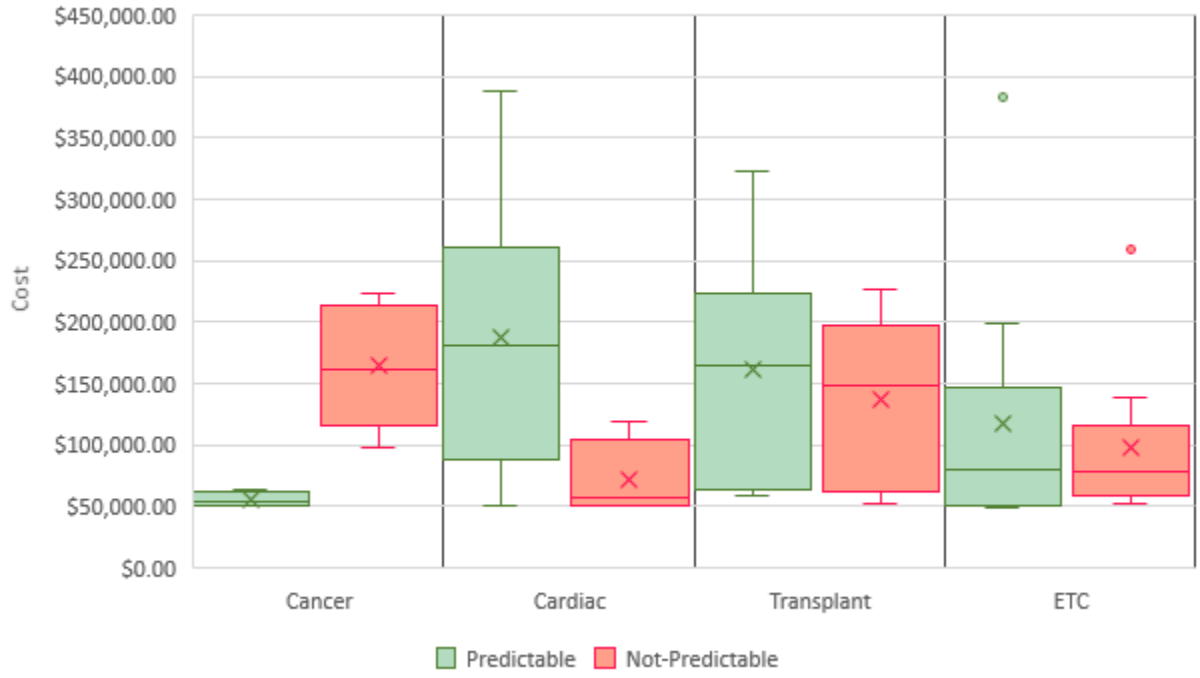


Figure 3.15 Comparison of costs between predictability at UWMC/Harborview

While for NW hospital, patients under the infection that were not predictable did have a slight higher cost range compared to those that were predictable. Despite the low frequency of non-predictable patients for both chronic and ETC, they still show that non-predictable costs for both of those categories are lower than their counterparts.

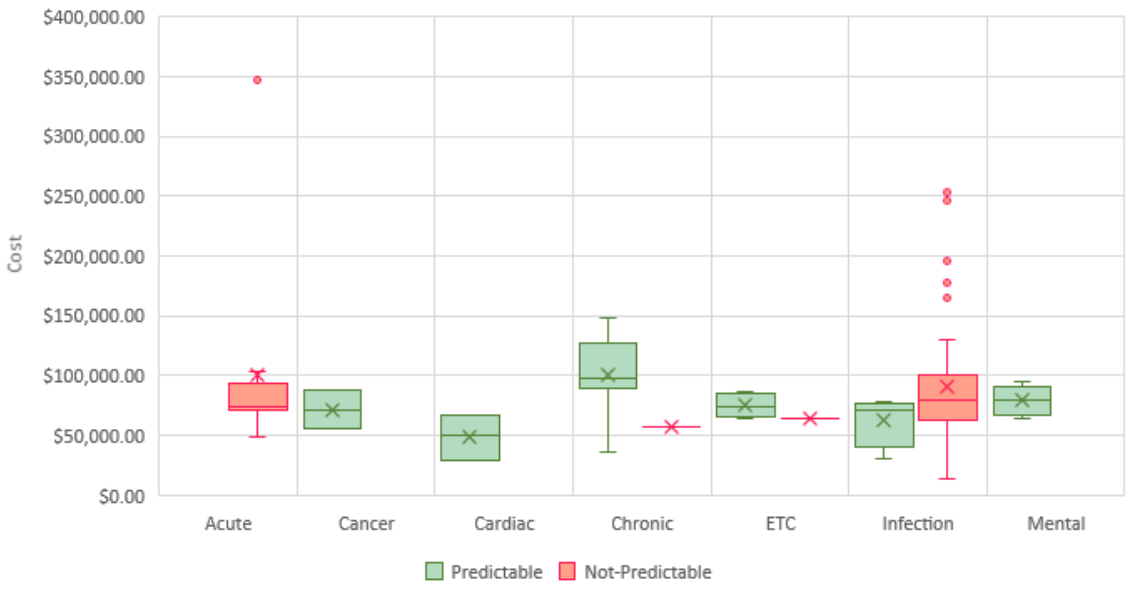


Figure 3.16 Comparisons of costs among categories between hospitals at NW Hospital

Discussion:

How much utility can EHR provide towards predicting high-cost patients? And to what extent can we use the predictive models generated?

Out of the 100 patients as part of the UWMC/Harborview cohort, only ⅓ of the patients I found to be predictable. To reiterate, predictability is based off of the root cause of the highest cost expenditure of the patient. This cost might be generated on one visit or across multiple visits at which point the root cause was take in tandem. 71 patients from UWMC/Harborview and 64 patients from Northwest Hospital were not predictable. The majority of the 71 patients were made up from acute and transfer patients. While the 64 patients comprised mostly out of infection patients. Out of the

predictable patients, a majority of them were under the cardiac and etc categories. While for NW hospital, patients are evenly spread without any sort of weight towards a category.

Looking at the makeup of the patients that weren't predictable, some clarity surfaces. Either their past medical history or cause of their high cost spending is naturally just unpredictable and there isn't anything to change that. Among our categories(acute, cardiac, transfer, transplant, and cancer) for UWMC/Harborview, acute and transfer are the categories that are significantly unpredictable. But between them they cover different aspects of why they aren't predictable. Patients with acute causes are naturally going to be unpredictable. These events such as gunshots, motor vehicle accidents, and burns are improbable to predict. They arise from random events that don't have a meaningful link with the prior health of the patient. Referencing the healthcare workforce, these patients would be considered episodic spending patients. (The et al., 2015) Transfer is a little bit more nuanced one why they can't be predictable. Instead of the unpredictable nature of the cause being more of a prevalent issue, it is because of the barrier to access the patient's history as well as the dynamic behind the transfer. In most cases when a patient is transferred, the patient has some diagnoses that is outside the ability of the current facility and thus the patient needs to be transported to a facility of greater capacities in order to efficiently treat the patient.(Kulshrestha & Singh, 2016) In the case of UWMC/Harborview, the hospital covers such a broad region

as the umbrella for WWAMI system(Washington, Wyoming, Alaska, Montana and Idaho). Given the breath of what they cover, it is quite understandable they receive many transfers. While going through patients, I found that mainly of them would have an abrupt start of their medical files and also mainly reside outside of the local region. When the patient gets transferred, they often start off with no prior visits because their past medical history is located in their local hospital instead and it takes time for any data to transfer over. Then getting the off-site data transferred and incorporated into UWMC would need to be approved by the attending physician. This happening after the fact (post cost incurrent). Thus any point of using the data to predict the high cost is defeated. This is the main reason on why there is such a high portion of transfer patients that are unpredictable. Not all transfers are unpredictable since the patient usually has previous medical records if they have already been admitted to the hospital. Now the question still stands that if we did have immediate access to other hospitals, could patients be predicted from further access. Fortunately, as I was going further into the EHR platform, I was able to access EPICaccess, which is web portal of the patient's previous medical record outside of UWMC that can be viewed from within the UWMC's platform. From there, I was able to see that there actually prior medical data on those patients, but it meant that the data wasn't actually in the UWMC database. Thus models currently can't use visual data not to mention the issues of privacy and interoperability. But if this extra set of data, it might make it more to make the high-cost

patient predictable. There is a logistical issue that and also the prevention would not be at the hands of UWMC. Instead the local hospital that the patient was being transferred to would most likely have the job of prevention. Overall the dynamics of transfer are inherently unpredictable and to make them predictable would likely require too many resources and conditions.

As far as the rest of the categories, transplants are mainly made up of non-predictable patients, but do cross over the predictable in cases. While I was looking at the patient records, I found that some of the transplants would have a referral going back months or a year before the actual high cost event. I realized that transplants have an interesting dynamic given the waitlists and the process in which to get the donor organ. Some transplant patients can either be part of the spectrum that gets rushed to the hospital in need of an organ. This usually is displayed in the medical records as some event record with little to no prior entries for the patient, which is the case for all of the non-predictable transfer patients. This makes predictability improbable for a model without further information. The alternative is that the transplant patient could have been waiting for the donor organ through a referral and have repeat appointments till the date of the surgery. This is what was shown in the few predictable transplant patients.

The last three categories are more predictable compared to the first two. With cancer patients, I was puzzled by what I found. Initially, I assumed that these patients would be predictable. There might not be anything to change the cost of the chemotherapy as it

is expensive inherently. But what I learned is that cancer isn't significantly predictive through the EHR. Many times, cancer patients are referred from outside the hospital. In this case, many patients were referred by the nearby Seattle cancer care alliance or from out of the state. This fact alone resonates back to the transfer patients, in which there is a block in the flow of data. So, these cancer patients wouldn't have any previous visit data except for when they come in for the chemotherapy treatments. Without any form of data to go off, then predictability becomes near improbable. Some other times there are cases were the cost is even higher than most, but this comes from an added complication of inflection, which is also seldom predictable.

Cardiac patients were harder to determine if they were predictable. Cardiac diseases at least from sifting through the records have many different diagnosis, degree of severity, and overlapping conditions. Also, as mentioned above hypertension doesn't have direct cost, but chronic heart failure does. (The et al., 2015) So navigating those nuances was hard. When I was looking at cardiac patients, they usually had a lot more visit records than that from other categories. This helped understand the timeline of what happened to the patient. Those that weren't predictable did run into the same issue as others with the lack in previous medical records. Given that UWMC/Harborview covers a large regions, many come to UWMC/Harborview for further treatment on their conditions or surgeries for severity. This manifests in many implantable devices to which many patients are referred over making it impossible to predict these.

Moving onto Northwest Hospital, the dynamic is completely different, and this is clearly shown in what is predictable and the addition of infection and mental categories. What surprised me was the infection category because it seemed so out of place. What was more astonishing was the prevalence of infection within the 100 patients at Northwest hospital.

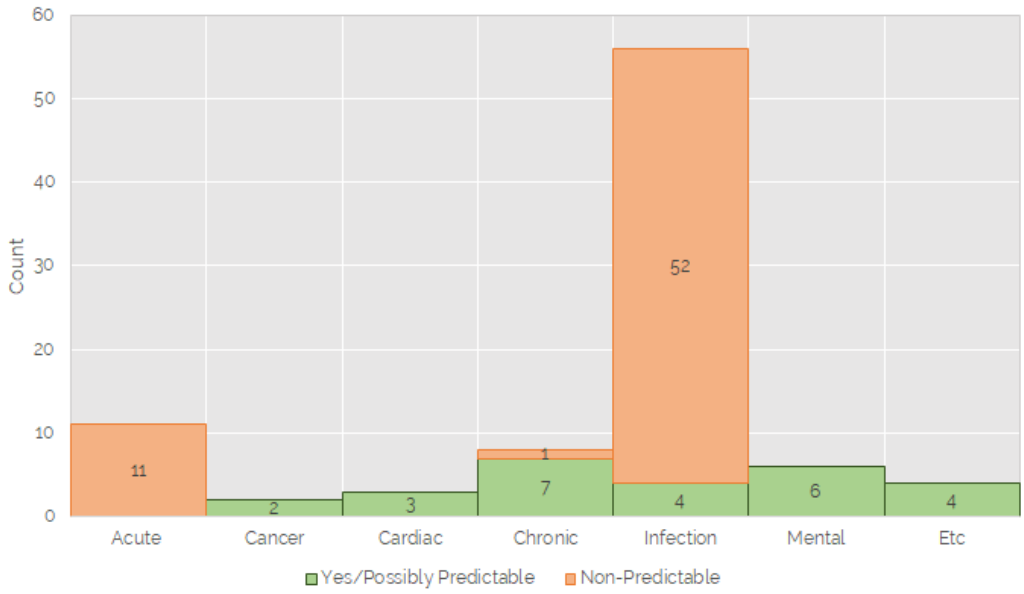


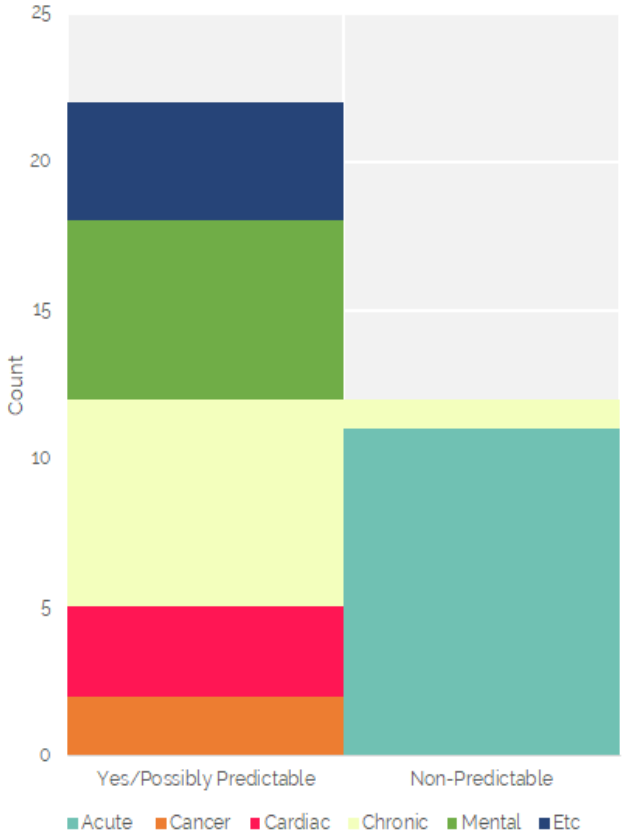
Figure 4.1 Ratio of unpredictable to predictable patients in categories at NW Hospital

As one can clearly see, infection outnumbers any other category. What categorizes infections includes both acute infections and sepsis. I don't have a definitive reason for why there are so many more cases, but I believe that given the lack of acute incidents as they would all go to UWMC/Harborview and NW Hospital is usually caters towards inpatient and outpatient, there are less abnormal cost incidents that can be so expensive besides infection. Patients with unpredictable infections usually had the most expensive

cost than any patients within the Northwest Hospital cohort. As for most of the other categories, they are mostly all predictive given just the amount of patient data that is available. This makes sense because as an inpatient/outpatient hospital the patients who are admitted tend to have recurring visits.

Originally, I previously thought that when I tallied the patients for predictability at NW Hospital, there would be more predictive patients than in relation to UWMC/Harborivew. I was wrong at face value, but if one removes infection from the frame, the trend becomes more of what I had imagined. There are more predictable patients in comparison. Purposefully eliminating the infection patients does impose a bias. But instead of seeing this change as underhanded manipulation of the data to get a desired result, it does show that there is some level of increase in predictive power when looking at patients that are common throughout both facilities.

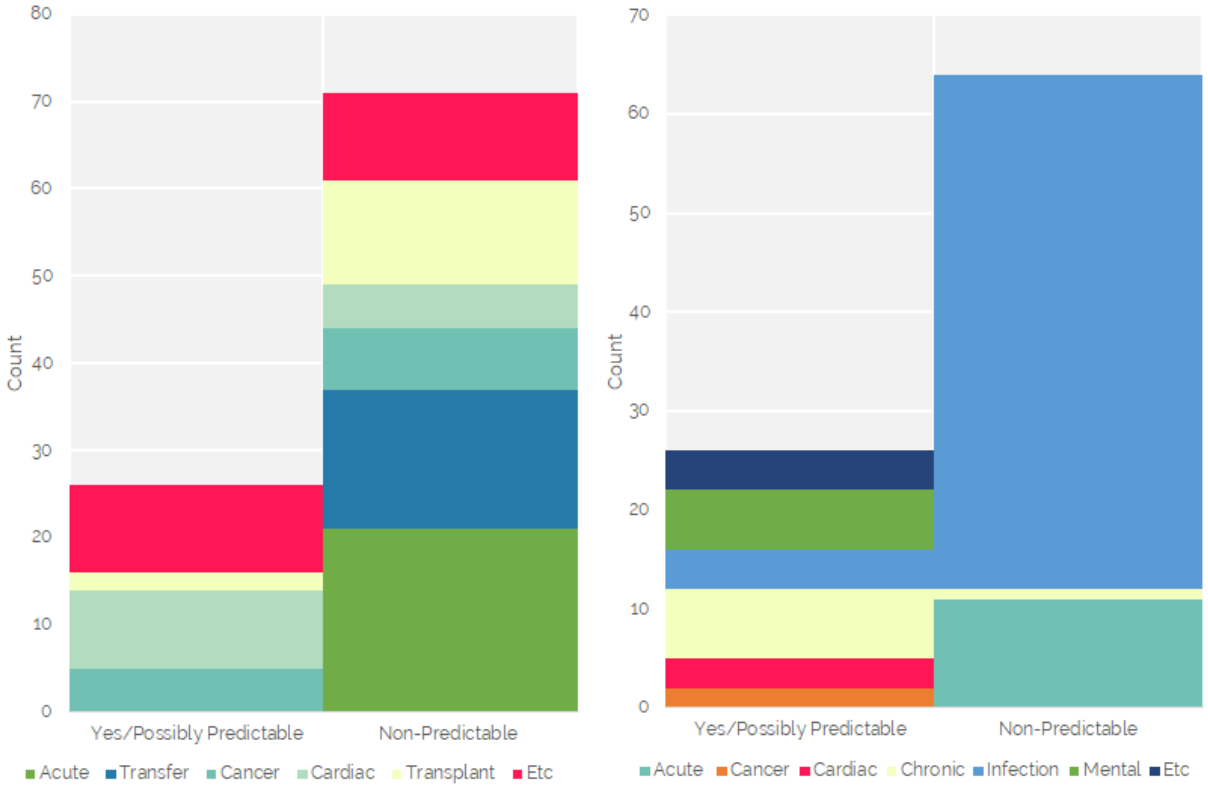
Comparisons to models:



Comparing what I found to what models usually predict, there seems to be a difference. Most tout that they have a predictably of 50% or more, while what I find is that majority of high-cost patients aren't predictable. I can't say for certain why there is such a noticeable difference. One theory is that models take more data than any one hospital has and so they can make more predictions. And this does make some sense given that if the non predictable patients in my cohort had all of their medical records available in the UWMC/Harborview/NHW database then their predictably would probably increase. Yet there is a silver lining in that not all hospitals would have the same ability to collaborate on such an interoperable model as the ones that have achieved such a high accuracy. Many of the regions that patients who had

missing prior EHR records in the database originated from more rural regions and probably don't have the capacities to fully comply with hospital data networking.

It was interesting to see how different the types of high-cost patients varied between the two facilities. I knew they would be different, since UWMC/Harborview is a hospital of a higher capacity that can take care of a broader and more intensive patients, their high-cost patients will reflect that. While the inpatient/outpatient community oriented



hospital of Northwest, would predominantly be more defined by chronic conditions. So in my initial thoughts, I presumed that Northwest would have a higher frequency in their cardiac and chronic patients. Yet this was completely different than what was the actual breakdown with infection being extremely high in frequency. This simply shows

how unique each institution high-cost patients nature's are and possibly the reason why models work for that particular institution or region is that the models pick up on that regions unique high cost nature, but when applied outside, then the model doesn't perform as well. An example of this was the systematic study on readmission models in which they reproduced numerous models of touted decent accuracy only to find that the accuracy is poor in most models. (Kansagara et al., 2011) Models that do perform better are only applicable in certain situations and setting. This might explain why such trends happen. Coming back to why there isn't such a main model that predicts high-cost patients, models aren't robust enough for such widespread and diverse environments. There are nuances in the patient clinical notes that aren't available in structured data in the EHR. Being able to use a human mind and a validation of a physician expertise was understanding that these patients were or were not predictable achieved. Assuming that NLP becomes so good that it can mimic how the human mind can understand health literature, still results in such a low predictive utility.

Cost understanding:

Looking through the cost data, I felt it important to see how varying the cost of each type of high-cost patients were. The clearest distinction was how tight the cost ranges of all categories at NW Hospital was in comparison to UWMC/Harborview. Comparing the acute costs for both facilities there seems to be some form of constriction for NW Hospital. NW Hospital gets more of a consistent degree of patients relatively and as a

reflect the high expenditure maintains a level of consistency. Compared to the more spread out high cost at UWMC might mean that the high cost isn't as consistent as the variety of conditions for each high-cost patient incident may vary.

Limitation:

There are inherent limitations in this study. The main limitation would be size of the sample coherent. This study examines patient data in order to better judge the utility of models that have thousands to millions of patient data. In comparison, having only 200 patients doesn't have the same level of power that the observations made would reflect well in a generalized manner. This limitation is set physically given that looking through 200 patients manually is time consuming and can't be expedited or scaled immensely in contrast if structured data was used instead. Another limitation is the fact that these observations were made through subjective qualitative measurements in clinical notes. They were verified by an external physician for reliability, but there is still a degree of subjectivity to the measurements. Thus, biases involved. Again, because there was not really any form of hard quantitative data my results lack concrete statistical analytics. So there is not form of definitive correlation or significance that can be claimed.

Additional Studies

If given more time and resources, I would expand the number of patients that I could analyze as well as the number of facilities to draw from. UWMC/Harborview and NW Hospital are both distinct in how they run, which did help give differing observations, but they still are under the same umbrella organization. I would have liked to taken hospitals outside of the local region, but still in the same umbrella, such as the Spokane campus to understand ways of minimizing the information wall so analysis can be done more smoothly despite the distance. It can be hard already to obtain access outside of one hospital but getting data access outside of hospital networks is even harder challenge. Branching out of network hospitals such as Swedish Medical Centers, Overlake medical Center and Providence Hospitals, would expand the assessment on how much barrier toward access to data currently in place. Another question I want to pose is what methods are needed to overcome the lack of interoperability among different hospital systems so that models don't need to be in the same system in order to achieve uniformity in their data to make a model feasible. Another expansion with the results would be to look into way of help models not predict high-cost patients that are improbable to prevent incurring of their high cost. As seen in transfer patients, if a model is able to predict these patients, it takes away

clinical impact from the models results as there isn't any way to improve upon the inherent high cost of transfers generally. But by filtering out patients that despite predictably, lack ways of prevention would then increase the impact of models.

Conclusion:

Current models that predict high-cost patients claim to have a certain predictive value, but how reliable and reproducible are they? There are countless models created touting some accuracy in predicting high costs among a large patient cohort. The creation of such a model might simply have added more to the complexity (in the form of more data or new methods) in the hopes that this variation will increase the accuracy. While the accuracy does increase, it doesn't seem to be transferable and reproducible outside of the facilities were the model was created. The fact remains that none of these models have persisted and have been used in a widespread manner, simply calls into question if models are approaching high-cost patients in the most effective method possible for prediction. There needs to be more done in the area of understanding why patients, in fact, become high cost in the most natural state of the EHR.

In this study, I evaluated 200 patients (100 in UWMC/Harborview and 100 in Northwest Hospital). The aim is through assessing each patient's predictability the reason behind their high cost becomes clear. Through this assessment, I hope that I can

also understand why models aren't as easily transferable across facilities and in a practical clinical setting.

After data collection, around ⅔ of all patients in both cohorts were not predictable. Categories that were completely unpredictable involved acute and transfer patients with some ratio of predictably in transplant, cardiac, and cancer. The unpredictably of these patients although evaluated through qualitative methods still have direct implications towards quantitative models. I discovered that one of the root causes for unpredictability stemmed from the nature of the event itself. Namely the lack of necessary prior patient data. Many times there would be no health records prior to the high-cost visit incurnment. This means that no model would have any chance at a predictive value given the sparse amount to lack of patient data available. This definitely is a hallmark of patients with acute injuries, but as mentioned in the discussion this lack of prior data goes beyond just acute patients. The contrast between what models claim compared to the findings in the UWMC patient cohort may highlight how unrealistic the conditions are for predictability. Models currently try to incorporate more and more data with even more features as a way to increase accuracy, but this may not be practical. Even if there are prior medical records, they are stuck behind independent databases from out of network hospitals.

Out of many of the categories, cancer, cardiac, and transplant are the types of categories that are inherently expensive and are more predictable. They often have

more visit data and often have visits that help with making high costs predictable. Although even the cost can be predictable, there isn't much that can be done to prevent such costs. Chemotherapy for cancer, implantation surgeries, and required expensive surgeries for transplants are all unavoidable costs. Thus prediction of these patients is not always completely clinically impactful.

The next finding was between large hospitals and smaller community hospitals. When the assessment of predictability of high cost is applied to smaller hospitals, there were significant differences compared to UWMC/Harborview. In Northwest Hospital, the largest category became infection. But there was such a huge weight towards the frequency of infection high-cost patients. This motif from NW Hospital shows how dissimilar costs from one hospital to another can be and a possible explanation towards why models don't fair well outside the place of origin. Each hospital may have their own different high cost dynamic that the model can not adapt to.

The streetlight effect is apt analog on the state of predictive high cost modeling. Current improvements simple go for the easiest area to look: more data. But my study shows, predictability is intricate and involve a further understanding of the structure and dynamics of the patient's data to truly understand the high cost nature. Models moving forward should instead of looking at what is the easiest approach, look at what is the true cause of the high cost in order to provide the most reliable models. The might change as more thought is put into methods for modeling. As machine learning and

feature selection become more adept in fine tuning to the patient data, they become more knowledgeable in the mechanisms of predictably. The helps models come closer towards being suited towards real world environments.

References

Ash, A. S., Ellis, R. P., Pope, G. C., Ayanian, J. Z., Bates, D. W., Burstin, H., … Yu, W. (2000). Using diagnoses to describe populations and predict costs. *Health Care Financing Review*, *21*(3), 7–28. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11481769

Berk, M. L., & Monheit, A. C. (1992). The Concentration of Health Expenditures: An Update. *Health Affairs*, *11*(4), 145–149. https://doi.org/10.1377/hlthaff.11.4.145

Billings, J., Dixon, J., Mijanovich, T., & Wennberg, D. (2006). Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. *BMJ (Clinical Research Ed.)*, *333*(7563), 327. https://doi.org/10.1136/bmj.38870.657917.AE

Calver, J., Brameld, K. J., Preen, D. B., Alexia, S. J., Boldy, D. P., & McCaul, K. A. (2006). High-cost users of hospital beds in Western Australia: a population-based record linkage study. *The Medical Journal of Australia*, *184*(8), 393–397. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16618238

Chechulin, Y., Nazerian, A., Rais, S., & Malikov, K. (2014). Predicting patients with high risk of becoming high-cost healthcare users in Ontario (Canada). *Healthcare Policy = Politiques de Sante*, *9*(3), 68–79. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24726075

Cohen, S. (2012). The Concentration of Health Care Expenditures and Related Expenses for Costly Medical Conditions, 2009. *Medical Expenditure Panel Survey*, *236*(February), 1–9. Retrieved from https://meps.ahrq.gov/data_files/publications/st455/stat455.pdf

Components of Indirect Costs. (n.d.). Retrieved June 4, 2018, from https://www.hopkinsmedicine.org/research/resources/offices-policies/ora/handbook/appendixd.html

Fleishman, J. A., & Cohen, J. W. (2010). Using Information on Clinical Conditions to Predict High-Cost Patients. *Health Services Research*, *45*(2), 532–552. https://doi.org/10.1111/j.1475-6773.2009.01080.x

Ghassemi, M., Celi, L. A., & Stone, D. J. (n.d.). State of the art review: the data revolution in critical care. https://doi.org/10.1186/s13054-015-0801-4

Greenberg, J. (2016). Calibrating the "Hot-Spotting" Hype. *MEDPAGETODAY*. Retrieved from https://www.medpagetoday.com/blogs/disruptions-or-distractions/60380

Hayes, S. L., Salzberg, C. A., Mccarthy, D., Radley, D. C., Abrams, M. K., Shah, T., & Anderson, G. F. (2016). High-Need, High-Cost Patients: Who Are They and How Do They Use Health Care? A Population-Based Comparison of Demographics, Health Care Use, and Expenditures. Retrieved from http://www.commonwealthfund.org/publications/issue-briefs/2016/aug/high-need-high-cost-patients-meps1

Health Human Services. (2016). Medical Expenditure Panel Survey Background. Retrieved June 4, 2018, from https://meps.ahrq.gov/mepsweb/about_meps/survey_back.jsp

Hong, C. S., Siegel, A. L., & Ferris, T. G. (n.d.). Caring for High-Need, High-Cost Patients: What Makes for a Successful Care Management Program? Retrieved from http://graceteamcare.indiana.edu/content/Care Management Complex High Cost Hong TCF 2014 (2).pdf

Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., & Kripalani,

S. (2011, October 19). Risk prediction models for hospital readmission: A systematic review. *JAMA - Journal of the American Medical Association*. American Medical Association. https://doi.org/10.1001/jama.2011.1515

Kulshrestha, A., & Singh, J. (2016). Inter-hospital and intra-hospital patient transfer: Recent concepts. *Indian Journal of Anaesthesia*, *60*(7), 451–457. https://doi.org/10.4103/0019-5049.186012

Lu, J., Britton, E., Ferrance, J., Rice, E., Kuzel, A., & Dow, A. (2015). Identifying Future High Cost Individuals within an Intermediate Cost Population. *Quality in Primary Care*, *23*(6), 318–326. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/27212892

Miao, Y., Qian, D., Sandeep, S., Ye, T., Niu, Y., Hu, D., & Zhang, L. (2017). Exploring the characteristics of the high-cost population from the family perspective: a cross-sectional study in Jiangsu Province, China. *BMJ Open*, *7*(11), e017185. https://doi.org/10.1136/bmjopen-2017-017185

Moturu, S. T., Johnson, W. G., & Liu, H. (2010). Predictive risk modelling for forecasting high-cost patients: a real-world application using Medicaid data Predictive risk modelling for forecasting high-cost patients. *Int. J. Biomedical Engineering and Technology*, *3*(12), 114–132. Retrieved from https://pdfs.semanticscholar.org/634a/2c81b378848bdfb2b020b2dd0608c2f26cd4.pdf

NIHCM Foundation. (2012). *The Concentration of Health Care Spending*. Retrieved from https://www.nihcm.org/topics/cost-quality/concentration-of-us-health-care-spending

Predicting high cost patients - Semantic Scholar. (n.d.). Retrieved June 4, 2018, from https://www.semanticscholar.org/search?q=predicting high cost patients&sort=relevance

Smith, S., Brick, A., O'Hara, S., & Normand, C. (2014). Evidence on the cost and cost-effectiveness of palliative care: A literature review. *Palliative Medicine*, *28*(2), 130–150. https://doi.org/10.1177/0269216313493466

The, I. F., Care, H., & Task, T. (2015). Proactively Identifying the High Cost Population. *Health Care Transformation Task Force*, (July). Retrieved from https://hcttf.org/wp-content/uploads/2018/01/WhitePaper-ProactivelyIdentifyingtheHighCostPopulation.pdf

Weiss, A. J., & Elixhauser, A. (2012). Overview of Hospital Stays in the United States, 2012. Retrieved from https://www.hcup-us.ahrq.gov/reports/statbriefs/sb180-Hospitalizations-United-States-2012.pdf